

~~SECRET~~

```

/db_xref="GI:2665458"
/translation="MSSPKRYGKGTGRNDDEEGNSGNTLNLSLPSSSCKKPTPTKPT

```

/translation="MSSPRKVRGKTGRNDEEEEGNSGNTLNRLNSLPSSSQKMTPTKPI

FMKNSKNSKPSHLSFSDKLYEYLPBLESDETDVSDRSLSNBNNSDSETA
GPKSTVSPVLSGSSSLSKDLETSEKELTOLAIPLFENSSARSMBRLTLOK
EELDIIEBIELOMPLPDDPNSGANTLOMRDRIIDLPYDSTRPLRGKRYXNS
YIRIVNHEEYVYIAOGLPEPIEDPMOMLENENVMATREIECQIKYKSYR
ISLKEPLREHVSYLETHHTVOTFYVAFQIVKSTGSOCVHLQETKMPHGTFA
SADFEIKYVRYKSHIPLVHCAGSGVGFICVDVFSALKNSFSDIMNIVT
OMRORCMIOREJOYOFCEYIEVLEVLQMLALY*

ASE COUNT 904 a 603 c 647 g 936 t

RIGIN

Query Match 16.8% Score 308; DB 10; Length 3090;
Best Local Similarity 59.3%; Pred. No. 3, 3e-40;
Matches 1079; Conservative 0; Mismatches 575; Indels 166; Gaps 26;

102 CAGATACGAGATATCTTCATTCATCATCATATAGTGGCCCAATAGAGAAC 161
1347 CTGCCATGAGAGAACTCTTTGACATTAAGACATAGACCAATAGAGAAC 1406
162 AACGTTGCGATGTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 221
1407 AGCGGTGCGATGTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1466
222 AAGTCTTGGAACTCTGACTTGGATTAAGAAAGACTCTGCTGCTGCTGAA 281
1467 AAGTCTTGGAACTCTGACTTGGATTAAGAAAGACTCTGCTGCTGCTGAA 1525
282 ATTACCAAGTGGCTTGCACCTCCATTAAGAAAGACTCTGCTGCTGCTGAA 341
1356 GTTACCGAGACCTGAGAGCC-----TGAGCGGTGCTGAGAGCC-TC 1565
342 TGCATGACATCAATCTGCTTCTTCTTATCAGTTATTTCTTCTTCTTCTT 401
1566 TGGGGGGGGTGGAGCTGCTC-----TGATTTTCTCTGAGAGCTCCT 1612
402 GAGGGGCAATATCATTTGGCTTGGGATCATGTTTACTATGATCTTCTGAGCA 461
1613 GAGGTGACACTACTGGGACAGAGTGAACCTGTTCCAC---TTGATCTTCTGAGCA 1669
462 TATCAAAATTAATCTCCCATTTTCCAGTGAACATGTTACTATAAAGATTGACGT 521
1670 GAGCAAAATTAATCTCCCATTTTCCAGTGAACATGTTACTATAAAGATTGACGT 1728
522 TGGCATTTGTTGAAGGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 581
1729 TGGCATTTGTTGAAGGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1787
582 AATTTATTTGGAAGGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 641
1788 AATTTATTTGGAAGGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1846
642 TTTGGGAG 698
1847 CTCTGGGAG 1906
699 T--GCTATTCATGTTCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 756
1907 TTGGAG 1963
757 GTTAAAAAG 816
1964 --TCAAAAG 2021
817 ACATTTTATA-----CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 842
2022 ATACTTCATGTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2081
843 AGTGTGAG 901
2082 AGCAGTCAAG 2141
902 -----TCGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 942
2142 TTAATATCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2201

943 AGCCAG 998
2202 ATTC--AGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2259
999 TACATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1058
2260 TACATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2319
1059 CTTCGCCCTTACGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1118
2320 CTTCAG 2373
1119 TTCAATGAG 1175
2374 TTTTGGTACCTGTTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2433
1176 TT-----TCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1230
2434 TTTTTCCTTACAG 2479
1231 CATTAG 1296
2480 TATTAG 2538
1291 GGAAG 1350
2539 AAAATTTCCAG 2596
1351 GAGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1410
2597 -----GAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2469
1411 TACCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1464
2650 CTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2709
1465 AATGAG 1524
2710 TGAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2769
1525 AATTTCTGAG 1584
2770 TGGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2829
1585 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1643
2830 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2889
1644 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1699
2890 AAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2949
1693 AATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1752
2950 ATATCATTTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3009
1753 CTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1812
3010 CTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3069
1813 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1832
3070 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3089

RESULT 8
AC130106 173224 bp DNA linear HTG 08-AUG-2002
LOCUS
DEFINITION
Rattus norvegicus clone CH230-17N23, *** SEQUENCING IN PROGRESS
AC130106
AC130106.1 GI:22138328

```

DR PROSITE; PSS0055; TYR_PHOSPHATASE_PTF; 1.
KW SIGNAL.
FT SIGNAL.
FT NON_TER
SQ SEQUENCE 297 AA; 34221 MW; BCFAA633BEF526DF CRC64;

Query Match
Best Local Similarity 24.0%; Score 88.5; DB 11; Length 297;
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3

Db 136 EEFNPSLPSGHIOGTFELANKEENREKNRYNIILPNDCRVLLSOLDGIPCSDYINASYID 195
      :||| : ||| : | : ||||| ||| |
      40 GY-----SGPNERTT--FW 51
      || || | |
      196 GYKERNKFIAAGCPKOETVNDWF 218

RESULT 4
O8TE48 PRELIMINARY; PRT; 536 AA.
ID O8TE48 AC O8TE48;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Tyrosine phosphatase epsilon pD1.
DN PRPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wabacken T., Hauge H., Funderud S., Aashelm H.C.;
RT "Characterisation, expression and functional aspects of a novel
RL protein tyrosine phosphatase epsilon isoform."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ30580; CAD3382.1;
SQ SEQUENCE 536 AA; 60799 MW; 06C29CBACBCCEBA CRC64;

Query Match
Best Local Similarity 24.0%; Score 88.5; DB 4; Length 536;
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

Db 2 QEFMALELNLPGEFNNGNOPSNREKNRYDILPQH-----H 39
      :||| : ||| : | : ||||| ||| |
      79 EEFNPSLPSGHIOGTFELANKEENREKNRYNIILPNDCRVLLSOLDGIPCSDYINASYID 136
      || || | |
      40 GY-----SGPNERTT--FW 51
      || || | |
      139 GYKERNKFIAAGCPKOETVNDWF 161

RESULT 5
O96K06 PRELIMINARY; PRT; 642 AA.
ID O96K06 AC O96K06;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE Tyrosine phosphatase epsilon.
DN PRPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wabacken T.K., Hauge H., Finne E.F., Miedeloch A., Aashelm H.C.;
RT "Expression of human protein tyrosine phosphatase epsilon,

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: June 11, 2003, 11:44:37 ; Search time 211 Seconds

(without alignments)
715.090 Million cell updates/sec

Title: US-09-935-703-7

Perfect score: 1 MOEFMALEKMLPGFENSGN.....TTFMHSNGCANVSLLRCA 67

Sequence: 369

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ .p2n .model -DEV=rlp
-Q=/cgn2_1/USPTO.spool/US09935703/runat_06062003.095210.6144/app_query.fasta.1.263
-DB=N-Geneseq.101002 -QFMT=fastp -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdl
-LIST=45 -LOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09935703.ecgn.1.1.396 -tunat_06062003.095210.6144 -NCP=6 -ICPU=3
-NO_MMAP -IARGSOERY -NEG.SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N-Geneseq.101002.*

- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	369	100.0	1832	22	AAH43389	PTPase cDNA clone
2	239.5	64.9	1914	24	AAH24022	Human protein phos
3	231.5	62.7	3264	23	AAH70229	DNA encoding novel
4	123	33.3	1692	20	AAH81746	Mouse PTP05 isofor
5	123	33.3	1785	20	AAH81744	Mouse PTP05 isofor
6	123	33.3	1896	20	AAH81745	Mouse PTP05 isofor
7	88.5	24.0	2160	24	ABK84570	Human cDNA differe
8	88.5	24.0	2389	22	ABK84570	Human cDNA differe
9	88.5	24.0	2393	22	AAH59773	Human polynucleoti
10	88.5	24.0	2689	22	AAH98486	Human polynucleoti
11	88.5	24.0	5425	24	ABH99311	Human EST-derived
12	88	23.8	1747	17	AAH35308	Mouse ischaemic co
13	86	23.3	1624	17	AAH35310	Rat M1PFP gene cDN
14	86	23.3	1788	22	AAH26454	Human SH-PTP1 cDNA
15	86	23.3	1788	22	AAH26456	SHP-1 DNA. Uniden
16	86	23.3	1788	22	AAH26467	SHP-1 activated do
17	86	23.3	1788	22	AAH26468	SHP-1 activated do
18	86	23.3	1788	22	AAH26469	SHP-1 activated do
19	86	23.3	1788	22	AAH26470	SHP-1 activated do
20	86	23.3	1788	22	AAH26471	SHP-1 activated tr
21	86	23.3	1788	22	AAH26472	SHP-1 activated tr
22	86	23.3	1788	22	AAH26473	SHP-1 activated qu
23	86	23.3	1788	22	AAH26474	SHP-1 activated qu
24	86	23.3	1788	22	AAH26475	SHP-1 activated qu
25	86	23.3	1788	22	AAH26484	SHP-1 activated qu
26	86	23.3	1788	22	AAH26485	SHP-1 E74A mutant
27	86	23.3	1866	21	AAH32917	Human HCP phosphid
28	86	23.3	2076	24	ABK34867	Human cDNA encodin
29	86	23.3	2145	17	AAH35306	Human SH-PTP1 gene
30	86	23.3	2175	24	ABK63571	Rat sequence differe
31	86	23.3	2277	21	AAH09637	Human Src homology
32	86	23.3	2277	21	ABK84611	Human cDNA differe
33	86	23.3	2315	24	ABO55031	Human ovarian anti
34	86	23.3	2475	21	AAH32916	Human HCP/GST fusi
35	86	23.3	2532	19	AAH71075	Green fluorescent
36	86	23.3	2562	22	AAH71076	Protein tyrosine p
37	81.5	22.1	14738	22	AAH02949	Human reproductive
38	81.5	22.1	14738	22	AAH02951	Human reproductive
39	81.5	22.1	222930	14	ABK84349	Human cDNA differe
40	79.5	21.5	2409	20	AAH36882	Human receptor typ
41	79.5	21.5	2409	20	AAH36882	Human receptor typ
42	79.5	21.5	2826	24	ABO55083	Human ovarian anti
43	79.5	21.5	2872	13	AAH021001	Marine receptor ty
44	79.5	21.5	2872	17	AAH36881	Marine receptor ty
45	79.5	21.5	2872	20	AAH29853	Mouse receptor typ

ALIGNMENTS

RESULT 1
ID AAH43389 standard; cDNA; 1832 BP.
XX
AC AAH43389;
XX
DT 15-NOV-2001 (first entry)
XX
DE PTPase cDNA clone HATBM23.
XX
KW Protein tyrosine phosphatase; PTPase; neural disorder; preservative;
KW Alzheimer's disease; Parkinson's disease; immune system disorder;
KW rheumatoid arthritis; Grave's disease; muscular disorder; wound healing;
KW reproductive disorder; pulmonary disorder; cardiovascular disorder;
KW arylhydantoin; infectious disease; viral hepatitis; multiple sclerosis;
KW hyperproliferative disorder; neoplasm; epithelial cell proliferation;
KW endocrine disorder; diabetes mellitus; transplantation; weight disorder;
KW hair loss; skin aging; sunburn; food additive; ss.

XX OS Homo sapiens.
 XX XX
 XX Key Location/Qualifiers
 XX FT 19..222
 XX FT CDS /tag= a
 XX FT /product= "Prtase"
 XX
 XX MO20016706-A1.
 XX
 XX PD 13-SEP-2001.
 XX
 XX PF 22-FEB-2001; 2001MO-US05496.
 XX
 XX PR 03-MAR-2000; 2000US-0186658.
 XX PR 16-MAR-2000; 2000US-0189881.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI
 XX PI Shl Y, Ruben SM;
 XX
 XX MPI; 2001-565582/63.
 XX P-PSDB; AAB47496.
 XX
 XX Novel isolated protein tyrosine phosphatase polypeptide useful for
 XX treating and preventing Alzheimer's disease, rheumatoid arthritis,
 XX Grave's disease, arrhythmias, neoplasms, multiple sclerosis and
 XX diabetes mellitus
 XX
 XX PS Claim 1; Page 276-77; 286pp; English.
 XX
 XX The sequences given in AAH43389-93 encode protein tyrosine phosphatase
 XX (Prtase) polypeptides. The Prtase polypeptides, or the cDNA encoding
 XX them, are useful for preventing, treating or ameliorating a medical
 XX condition in a mammalian subject. They are useful for diagnosing,
 XX preventing or treating neural disorders (e.g. Alzheimer's disease,
 XX Parkinson's disease), immune system disorders (e.g. Rheumatoid
 XX arthritis, Grave's disease), muscular disorder, reproductive
 XX disorders, pulmonary disorders, cardiovascular disorders
 XX (e.g. arrhythmias), infectious diseases (e.g. viral hepatitis),
 XX hyperproliferative disorders (e.g. neoplasms), diseases of the cellular
 XX level (e.g. multiple sclerosis), endocrine disorders (e.g. diabetes
 XX mellitus), and wound healing and epithelial cell proliferation.
 XX They are also useful for preventing hair loss, to prevent skin aging
 XX due to sunburn, to maintain organs before transplantation, to modulate
 XX mammalian characteristics, to treat weight disorders, to change a
 XX mammal's mental or physical state, or as a food additive or
 XX preservative.
 XX
 XX SO Sequence 1832 BP; 560 A; 299 C; 321 G; 652 T; 0 other;
 XX
 XX Alignment Scores:
 XX Pred. No.: 1,94e-41 Length: 1832
 XX Score: 369.00 Matches: 67
 XX Percent Similarity: 100.008 Conservative: 0
 XX Best Local Similarity: 100.008 Mismatches: 0
 XX Query Match: 100.008 Indels: 0
 XX DB: 22 Gaps: 0
 XX
 XX US-09-935-703-7 (1-67) x AAH43389 (1-1832)
 XX
 XX 1 MetGlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20
 XX Db 19 ATGCAAGAAATTATGCTTTAGAACTTAAAGATCTGCTGTGATCAACTCTGGGAAAT 78
 XX Oy 21 GlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHisGly 40
 XX Db 79 CAACCAAGCAACGAGAAAAAACAAGTACCGAGATTTCTTCATTTCACACATCATGGA 138
 XX Oy 41 TyrSerGlyProAsnGluArgThrPheTyrPHisGlySerAsnGluGlyAlaValSer 60
 XX Db 139 TATAGTGGCCCAAAATGAGAGAACACGCTTCTGGCATGTTCAACGAAGGACGATATCA 198

Oy 61 LeuLeuLeuArgTyrCysAla 67
 Db 199 CTTTGTACGATATGTGCT 219
 RESULT 2
 AAD24022
 ID AAD24022 standard; CDNA: 1914 BP.
 XX
 XX AC AAD24022;
 XX
 XX DT 26-MAR-2002 (first entry)
 XX
 XX DE Human protein phosphatase-4 CDNA.
 XX
 XX KW Human; protein phosphatase; PP-4; immune system disorder; AIDS; allergy;
 XX KW neurological disorder; developmental disorder; Alzheimer's disease;
 XX KW cell proliferative disorder; Huntington's disease; arteriosclerosis;
 XX KW renal tubular acidosis; gonadal dysgenesis; cancer; adenocarcinoma;
 XX KW Leukemia; transgenic animal; gene therapy; ss.
 XX
 XX OS Homo sapiens.
 XX
 XX FH Key Location/Qualifiers
 XX FT 203..1465
 XX FT CDS /tag= a
 XX FT /product= "Protein phosphatase-4"
 XX
 XX PN MO200196546-A2.
 XX
 XX PD 20-DEC-2001.
 XX
 XX PF 14-JUN-2001; 2001MO-US19442.
 XX
 XX PR 16-JUN-2000; 2000US-212447P.
 XX PR 22-JUN-2000; 2000US-213746P.
 XX PR 29-JUN-2000; 2000US-215210P.
 XX PR 06-JUL-2000; 2000US-216529P.
 XX PR 12-JUL-2000; 2000US-218080P.
 XX PR 21-JUL-2000; 2000US-220117P.
 XX
 XX PA (INCYTE) INCYTE GENOMICS INC.
 XX
 XX PI Au-Young J, Baughn MR, Ding L, Elliott VS, Gandhi AR, Griffin JN;
 XX PI Hafalia A, Kearney L, Lee EA, Lu Y, Nguyen DB, Patterson C;
 XX PI Ramkumar J, Reddy R, Sanjanwala MS, Stewart EA, Tang TT;
 XX PI Thornton M, Tribouley CM, Walla NK, Yang J, Yao MG, Yue H;
 XX
 XX MPI; 2002-090206/12.
 XX P-PSDB; AAE14454.
 XX
 XX Novel polypeptide, useful for diagnosing, treating or preventing
 XX disorders of growth and development, immune system, neurological and
 XX cell proliferation diseases, comprises cancer protein phosphatase
 XX polypeptides
 XX
 XX Claim 5; Page 112-113; 116pp; English.
 XX
 XX The present sequence is human protein phosphatase (PP)-4 CDNA.
 XX PP polynucleotide and polypeptide are useful in the diagnosis,
 XX treatment and prevention of immune system disorders, neurological
 XX disorders, developmental disorders and cell proliferative disorders.
 XX Examples of immune system disorders include acquired immune deficiency
 XX syndrome (AIDS), severe combined immunodeficiency disease (SCID),
 XX adult respiratory distress syndrome, allergies, amyloidosis,
 XX anaemia, asthma, atherosclerosis, Crohn's disease, atopic
 XX dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,
 XX Graves' disease, multiple sclerosis, myasthenia gravis, myocardial or
 XX pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis,
 XX psoriasis, Reiter's syndrome, rheumatoid arthritis, Sjogren's syndrome,
 XX scleroderma, systemic sclerosis, trauma; neurological disorders include
 XX Alzheimer's disease, Huntington's disease, dementia, epilepsy,
 XX Parkinson's disease, mental retardation and other developmental
 XX disorders of central nervous system such as Down's syndrome, cerebral

Best Local Similarity: 25.908
Query Match: 62.748
DB: 23
Mismatches: 2
Indels: 185
Gaps: 1

US-09-935-703-7 (1-67) x AAV81746 (1-3264)

```

QY      1 MetGlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20
DB      901 ATTCGAGGATTTATGGCTTTAGAACTTGTGCTGGAGTTCTACTCTGGGAAT 960
QY      21 GlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuPro----- 35
DB      961 CACCAACCAACAGAGAGAAAAAAGAGATACCGAGATATCTTCCATATGATTCAACACGC 1020
QY      35 ----- 35
DB      1021 GTTCCTCTGGAAAAAGCAAGACTACATCATGCTAGTATATAGAACTACATTTGT 1080
QY      35 ----- 35
DB      1081 GGAGAGAGATTTTTTATATGCTACTCAAGACACTGCTGAGCAGCATAGATGACTTT 1140
QY      35 ----- 35
DB      1141 TGGCAATGCTGTGGAAAAATATTCAAATGTTATGCCATGATTAACAGAGATGGAA 1200
QY      35 ----- 35
DB      1201 GGTCGAATTATCAATGCTACCATTTACTGCCCCATTCTCTGAAGAGCATTTGGAATTG 1260
QY      35 ----- 35
DB      1261 AACAACCTCCGTGTATTCCTGGAGAACTACAGATCTCAATATTTGATCATTCGAAATG 1320
QY      35 ----- 35
DB      1321 TTTCAGATTGTGAGAGATGCCAGGGAAGTACCTCTGTAAACAGTTGACATTCAC 1380
QY      35 ----- 35
DB      1381 AAGTGGCAGACATGGCAGCTCTGCTCAGCAGATAGCTTCAATAAATATATTCGTTAT 1440
QY      35 ----- 35
DB      1441 GCAAGAGAGACCACTTACAGGCCATGGTTGCTACGCAATGCCGCGCATAGCGCGG 1500
QY      36 -----Phe 36
DB      1501 ACAGGGGTTCCTATGTGTGATGTGTGTCGTCGTCATCGTAAGGACCTGTTCTTC 1560
QY      37 GlnHisHisGlyTyrSerGlyProAsnGluArgThrPheThrPHisGlySerAsnGlu 56
DB      1561 AA-CATCATGTGATATAGTGGCCCAATGAGAGAACAGCTTCTGCGCATGTTCAACGAA 1619
QY      57 GlnAlaValSerLeuLeuLeuArgTyrCysAla 67
DB      1620 GGAGCAGATACACTTTTGTACGATATTGCT 1652

```

RESULT 4
AAV81746 standard; cDNA; 1692 BP.
AAV81746:

10-MAR-1999 (first entry)

Mouse PTP05 isoform #2 encoding cDNA.

PTP04; PTP05; PTP10; SAD; ALK-7; protein tyrosine phosphatase;
type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
neurodegenerative disease; neuronal survival; Alzheimer's disease;
Parkinson's disease; Huntington's disease; ss.

OS Mus sp.
XX Key Location/Qualifiers
FH Key 199.1413
FT CDS /tag=a
FT /note="no stop codon given"

MO9849317-A2.

05-NOV-1998.

27-APR-1998; 98MO-US08439.

23-OCT-1997; 97US-0063595.

28-APR-1997; 97US-0044428.

20-MAY-1997; 97US-0047222.

11-JUN-1997; 97US-0049477.

11-JUN-1997; 97US-0049756.

18-JUN-1997; 97US-0049914.

(SUGEN) SUGEN INC.

App H, Clary D, Courtneidge SA, Hui TH, Jallat B;

Markby D, Onrust S, Peles E, Plovman GD;

MP1; 1999-009434/01.

P-PSDB; AAW89251.

Claim 2; Page 148-149; 193pp; English.

The present invention describes isolated, enriched or purified nucleic acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The present sequence encodes mouse PTP05. The above proteins, other than ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify substances that modulate their activity (i.e. agonists and antagonists, including NBP) in vivo or in vitro. These substances are used to treat or prevent diseases associated with abnormal signal transduction pathways that involve the proteins, particularly cancer (e.g. leukaemia and lymphoma), while modulators of ALK-7 (which is a type I receptor serine/threonine kinase) are used to promote neuronal survival. CC particularly for treating Alzheimer's, Parkinson's or Huntington's diseases. Nucleic acid fragments of the polynucleotides encoding the CC proteins can be used as probes to identify and clone related sequences; CC to detect protein-encoded RNA; to generate transgenic animals and in CC gene therapy (optionally after mutation). Ab are used to determine the CC proteins.

Sequence 1692 BP; 507 A; 342 C; 386 G; 457 T; 0 other;

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

4.21e-07
123.00
66.008
52.008
33.338
20
Length: 1692
Matches: 26
Conservative: 7
Mismatch: 16
Indels: 1
Gaps: 0

US-09-935-703-7 (1-67) x AAV81746 (1-1692)

```

QY      1 MetGlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20
DB      694 ATCCGAGATTTTGGAGTTAGAACAAATGACTCTGCTCATGATTCATTTCTGGGAAT 753
QY      21 GlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGln-HisHisG1 40
DB      754 ACACAGACAGACAGAGATTAAGAACAGATACCGAGATATCTTCATATGATTAACACGCT 813
QY      40 YTYrSerGlyProAsnGluArgThr 49

```

DB 814 GTTCCTCTTGAAAAACAGACTAC 841

RESULT 5
AAV81744
ID AAV81744 standard; cDNA: 1785 BP.

XX AAV81744;
AC
XX 10-MAR-1999 (first entry)
DT
XX
DE Mouse PTP05 encoding cDNA.
XX
XX PTP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 199..1479
ET /tag= a

PN MO9849317-A2.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98MO-US08439.
XX
XX 23-OCT-1997; 97US-0063595.
PR 28-APR-1997; 97US-0044428.
PR 20-MAY-1997; 97US-0047222.
PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
XX
XX (SUGEN-) SUGEN INC.
XX
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
PI Markby D, Onrust S, Peles E, Plozman GD;
XX
XX WPI: 1999-009434/01.
DR P-PSDB: AAM89249.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
PS Claim 2; Page 147; 193pp; English.

XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence encodes mouse PTP05. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival.
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.

XX Sequence 1785 BP; 527 A; 361 C; 408 G; 489 T; 0 other;
SQ

Alignment Scores: 4.52e-07 Length: 1785
Pred. No.: 123.00 Matches: 26
Score: 66.008 Conservative: 7
Percent Similarity:

Best Local Similarity: 52.00% Mismatches: 16
Query Match: 33.33% Indels: 1
DB: 20 Gaps: 0

US-09-935-703-7 (1-67) x AAV81744 (1-1785)

QY 1 MetGInGluPhemeAlaLeuCluLeuysAsnLeuProGlyGluPhenSerGlyAsn 20
DB 694 ATCCGTGAGTTTGGAGTTAGAACAAATGACTGTGCTGATGACTCAATTCGGGAAT 753
QY 21 GlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGln-HisHisG1 40
DB 754 AACTACAGAAACAGATTAAGACAGATACGAGATATCTTCATATGATCAACACCT 813
QY 40 YTYrSerG1ProAsnGluArgThr 49
DB 814 GTTCCTCTTGAAAAACAGACTAC 841

RESULT 6
AAV81745
ID AAV81745 standard; cDNA: 1896 BP.
XX
AC AAV81745;
XX
DE 10-MAR-1999 (first entry)
XX
DE Mouse PTP05 isoform #1 encoding cDNA.
XX
XX PTP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 199..1590
ET /tag= a

PN MO9849317-A2.
XX
PD 05-NOV-1998.
XX
PF 27-APR-1998; 98MO-US08439.
XX
XX 23-OCT-1997; 97US-0063595.
PR 28-APR-1997; 97US-0044428.
PR 20-MAY-1997; 97US-0047222.
PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
XX
XX (SUGEN-) SUGEN INC.
XX
PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
PI Markby D, Onrust S, Peles E, Plozman GD;
XX
XX WPI: 1999-009434/01.
DR P-PSDB: AAM89250.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
PS Claim 2; Page 147-148; 193pp; English.

XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence encodes mouse PTP05. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction

CC pathways that involve the proteins, particularly cancer (e.g. Leukemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival.
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.

SO Sequence 1896 BP; 572 A; 372 C; 427 G; 525 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	1896
Score:	123.00	26
Percent Similarity:	66.00%	7
Best Local Similarity:	52.00%	16
Query Match:	33.33%	1
DB:	20	0
Gaps:		

US-09-935-703-7 (1-67) x AAW81745 (1-1896)

OY 1 MetGlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20
Db 805 ATCCGTGAGTTTGGAGTTGAGAACAAATGACTGCGCTGAGGACTTCAATCTGGGAAT 864
OY 21 GlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGln-HisHISG 40
Db 865 ACACATACAGAACAGATATAGAACAGATACACAGATATCTTCATATATCATACACCGT 924
OY 40 YTYrSerGlyProAsnGluArgThr 49
Db 925 GTTCCTCTTGAAAAAACAGGACTACA 952

RESULT 7

ABK84570
ID ABK84570 standard; cDNA: 2160 BP.

AC ABK84570;

DT 14-AUG-2002 (first entry)

DE Human cDNA differentially expressed in granulocytic cells #1141.

KW Human: ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.

OS Homo sapiens.

PN MO200228999-A2.

PD 11-APR-2002.

PF 03-OCT-2001; 2001MO-US30821.

PR 03-OCT-2000; 2000US-237189P.

PA (GENE-) GENE LOGIC INC.

PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;

DR MPI; 2002-435328/46.

PT Detecting granulocyte activation by detecting differential expression
PT of genes associated with granulocyte activation, which serves as
PT diagnostic markers that is useful for monitoring disease states and
PT drug toxicity

XX Claim 1; SEQ ID No 1141; 114pp: English.

XX The invention relates to detecting (M1) granulocyte (GC) activation
XX (GCA), by detecting the level of expression of gene(s) (Gs) identified by
XX DNA chip analysis as given in the specification, and comparing
XX the expression level to an expression level in an unactivated
XX GC, where differential expression of Gs is indicative of GCA.
XX Also included are modulating (M2) GA by contacting GC with an agent
XX that alters the expression of at least one gene in Gs; (2) screening (M3)
XX for an agent capable of modulating GCA or an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease using the
XX gene expression profile; (3) detecting (M4) an inflammation (especially
XX chronic) in a tissue, an allergic response in a subject, exposure of a
XX subject to a pathogen or sterile inflammatory disease, by detecting the
XX level of expression in a sample of the tissue of gene(s) from Gs, where
XX the level of expression of the gene is indicative of inflammation;
XX (4) treating (M5) an inflammation (especially chronic) or in a tissue,
XX an allergic response in a subject, exposure of a subject to a pathogen
XX or sterile inflammatory disease, by contacting a tissue having
XX inflammation with an agent that modulates the expression of gene(s)
XX from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
XX GCA preferably in an inflammation in a tissue; M4 is useful for
XX detecting an inflammation (especially chronic) in a tissue, an allergic
XX response in a subject, exposure of a subject to a pathogen or sterile
XX inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
XX glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
XX reperfusion injury, ARDS, adult respiratory distress syndrome,
XX inflammatory bowel disease, Crohn's disease, ulcerative colitis,
XX periodontal disease; also bacterial infection, viral infection,
XX parasitic infection, protozoal infection, fungal infection, and M5 is
XX useful for treating one of the above conditions. The present
XX sequence represents a gene differentially expressed in granulocytes.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2160 BP; 578 A; 584 C; 561 G; 437 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	2160
Score:	88.50	24
Percent Similarity:	37.35%	7
Best Local Similarity:	28.92%	19
Query Match:	23.98%	33
DB:	24	3
Gaps:		

US-09-935-703-7 (1-67) x ABK84570 (1-2160)

OY 2 GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln 21
Db 460 GAGGATTCACACTGATGCGATCTGACACATACAGAACTTTGAACTGGCAATATA 519
OY 22 ProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis----- 38
Db 520 GAAGAAACAGAGAAAAAAGATATCCACATCTCCCAATGACCAATCTAGAGTG 579
OY 39 -----His 39
Db 580 ATTCTGAGCCAACTGATGGAATTCCTGTTCAAGTACATCAATGCTTCTACATAGAT 639
OY 40 GlyTyr-----SerGlyProAsnGluArgThr----- 49
Db 640 GGTACAAAGACAGATATTAATTCATAGCAGCTCAAGTCCCAACAGAAACGGTTAAC 699
OY 50 ---PheTTP 51
Db 700 GACTTCTGG 708

RESULT 8

AA157987
 ID AA157987 standard; cDNA: 2389 BP.
 AC AA157987;
 DT 22-OCT-2001 (first entry)
 XX
 XX Human polynucleotide SEQ ID NO 190.
 DE
 KW Human; neurotrophic; immunosuppressant; cytoskeletal; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 XX Homo sapiens.
 OS
 XX WO200153312-A1.
 PN
 XX 26-JUL-2001.
 PD
 XX
 XX 26-DEC-2000; 2000WO-US34263.
 PF
 XX
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 DR WPI: 2001-442253/47.
 DR P-PSDB; AAM38831.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS
 PS Claim 1; SEQ ID NO 190; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotrophic,
 CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilization of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 SQ Sequence 2389 BP; 656 A; 612 C; 584 G; 537 T; 0 other;

US-09-935-703-7 (1-67) x AA157987 (1-2389)
 QY 2 GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln 21
 Db 460 GAGGAGTCAACATCATGTCATCTGACACATACAGCACTTGTGAACGTGGCAAAATAA 519
 QY 22 ProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis----- 38
 Db 520 GAAGAAACAGAGAAAAAACAGATATCCCAATCCTTCCCAATGACCATTTAGAGGTG 579
 QY 39 -----His 39
 Db 580 ATTCTGAGCCACTGATGATGATTCCTGTTCAAGCATCATCATTCCTTCTACATAGAT 639
 QY 40 GlyTyr-----SerGlyProAsnGluArgThr----- 49
 Db 640 GGTACAAAGAGAAATTAATTCATAGACAGCTCAGGTCCCAACAGAAACGGTTAAC 699
 QY 50 ---PheTyr 51
 Db 700 GACTTCTGG 708
 RESULT 9
 AA159773
 ID AA159773 standard; cDNA: 2393 BP.
 AC AA159773;
 XX
 XX 22-OCT-2001 (first entry)
 DT
 XX
 DE Human polynucleotide SEQ ID NO 3762.
 XX
 XX Human; neurotrophic; immunosuppressant; cytoskeletal; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 XX Homo sapiens.
 OS
 XX WO200153312-A1.
 PN
 XX 26-JUL-2001.
 PD
 XX
 XX 26-DEC-2000; 2000WO-US34263.
 PF
 XX
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 DR WPI: 2001-442253/47.
 DR P-PSDB; AAM40617.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS
 PS Claim 1; SEQ ID NO 3762; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotrophic,

CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
SQ	Sequence 2393 BP; 634 A; 655 C; 615 G; 489 T; 0 other;
Alignment Scores:	
Pred. No.:	0.0444
Score:	88.50
Percent Similarity:	37.35%
Best Local Similarity:	28.92%
Query Match:	23.98%
DB:	
	Length: 2393
	Matches: 24
	Conservative: 7
	Mismatches: 19
	Indels: 33
	Gaps: 3

PR 15-SEP-2000; 2000US-0663870.
 XX (HYSE-) HYSEQ INC.
 XX
 XX
 PI Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi Y;
 PI Cao Y, Dirmanc RA, Zhang J, Werhman T;
 XX
 XX WPI; 2001-476164/51.
 DR P-PSDB; AAM23827.
 XX
 XX
 PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 PI antibodies and research use -
 XX
 XX
 PS Claim 1; Page 432; 1275pp; English.
 XX
 XX The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, mackerel, yeast, bacteria, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a cDNA
 CC of the invention.
 XX
 XX Sequence 2689 BP; 713 A; 706 C; 685 G; 585 T; 0 other;

OY		2	GlncluphemetalaleuGluleuysasnleuProGlyglIupheansnserglyAsngln	21
Dd		460	GAGAGATTCAACTATTTCCCATCTTGGACACATACAAAGAACCTTTTGAACTGGCAATAA	519
OY		22	ProSerAsnArgJulysAsnArtyrArgaspileleuProPheglnHis	38
Dd		520	GAGAAAAACAGAGAAAAAACAGATATGCCAACATCCTTCCCANAAGCACATTCAGGNG	579
OY		39	-----His 39	
Dd		580	ATTCTGAGCCACTGTGATGAATTCCCTGTTCAGACTACATCAATGCTTCTACATAGAT	639
OY		40	GLTYTr-----SerGlyProAsnGluArgThr----	49
Dd		640	GGTTCACAAGAAGATTAATTCTATGACAGCTCAAAGTCCCAACAGGAAGCGTTAAC	699
OY		50	---PheTrp 51	
Dd		700	GACTTCGTGG 708	
RESULT 10				
AAH98486	ID	AAH98486	standard; cDNA; 2689 BP.	
XX	AC	AC	XX	
XX	XX	XX	XX	
XX	DE	DE	XX	
XX	DT	DT	XX	
XX	12-OCT-2001	(first entry)	XX	
Human EST-derived coding sequence SEQ ID NO: 343.				
Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;				
tomato; monkey; dog; sea urchin; expressed sequence tag; EST;				
diagnostics; forensic test; gene mapping; genetic disorder;				
biodiversity; gene therapy; nutrition; ss.				
Homo sapiens.				
XX	OS	OS	XX	
XX	PN	PN	XX	
XX	MO200154477-A2.		XX	
XX	02-AUG-2001.		XX	
XX	PD	PD	XX	
PF	25-JAN-2001; 2001MO-US02687.		PF	
XX	XX		XX	
PR	25-JAN-2000; 2000US-0491404.		PR	
PR	17-JUL-2000; 2000US-0617746.		PR	
R	03-AUG-2000; 2000US-0631451.		R	

Alignment Scores:		0.0518	Length:	2689
Pred. No.:	88.50	Matches:	24	
Score:	37.35%	Conservative:	7	
Percent Similarity:	28.92%	Mismatches:	19	
Best Local Similarity:	23.98%	Indels:	33	
Query Match:	22	Gaps:	3	

US-09-935-703-7 (1-67) x AAH98486 (1-2689)	
QY 2 GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln 21	::: : :
Db 756 GAGGAGTTCAACTCTATTGCCATCTGGACACATACAGAACTTTTAACTGGCAATMAA 815	: :
QY 22 ProSerAsnAlaGgGluLysAsnArGlyrArGaspIleLeuProPheGlnHs----- 38	
Db 816 GAAGAAAAACAGAAAAAAACAGATATCCCAACATCCTTCCCAATGACCATTTAGGGTG 875	
QY 39 -----Hls 39	
Db 876 ATTCTGAGCCAACTGGATGGAAATTCCTGTTGACACTACATCAATGCTTCTCATATGAT 935	
QY 40 GlyTyr-----SerGlyProAsnGluArGThrThr--- 49	: :
Db 936 GGTTACAAAGAGAGAAATAATTCTATAGCAGCTCAAGGTGCCAAACAGAAAAAGTTAAC 995	: :
QY 50 ---PheTrp 51	
Db 996 GACTTCTGG 1004	

RESULT 11
ABI99311
ABI99311 standard; cDNA: 5425 BP.
XX
AC ABI99311;
XX
DT 07-MAR-2002 (first entry)
XX
Mouse Ischaemic condition related cDNA sequence SEQ ID NO:171.
DE Mouse; Ischemia; compressive ischaemia; occlusive ischaemia;
XX vasospastic Ischemia; Ischaemic condition; Ischaemic disease; ss.
XX
Mus musculus.
OS
XX
MO200186188-A2.
XX

```

XX 22-NOV-2001.
XX PF 18-MAY-2001; 2001MO-JP04192.
XX PR 18-MAY-2000; 2000JP-0145977.
XX PA (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;
XX DR MPI; 2002-034733/04.
XX P-PSDB; ABB57081.
XX PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring
XX PT expression levels of particular genes defined in the specification or
XX PT genes - determining the expression profile of a gene group comprising these
XX PS Claim 2; Page 505-511; 2690pp; English.
XX XX
CC The present invention describes a method for examining ischemic
CC conditions, comprising measuring the expression levels of particular
CC genes (1) in a test sample or determining the expression profile of a
CC gene group in the sample comprising genes selected from (1). The method
CC is useful for examining the ischemic condition (e.g. compressive
CC ischemia, occlusive ischemia or vasospastic ischemia) by measuring
CC expression levels of particular genes (AB199202 to AB199912, encoding
CC the protein sequences in ABB57020 to ABB57374) or by determining the
CC expression profile of a gene group comprising these genes. The
CC expression levels or expression profiles produced by these genes are
CC used as an indicator when screening for ischemic condition-improving
CC drugs or therapeutics for ischemic diseases. AB199913 and AB199914
CC represent PCR primers for a mouse ischemic condition related sequence,
CC which are used in the exemplification of the present invention.
XX SQ Sequence 5425 BP; 1430 A; 1339 C; 1344 G; 1312 T; 0 other:

Alignment Scores:
Pred. No.:      0.132      Length:      5425
Score:          88.50      Matches:     24
Percent Similarity: 37.35%   Conservative: 7
Best local Similarity: 28.92% Mismatches:    19
Query Match:      23.98%    Indels:       33
Db:               24        Gaps:         3

US-09-935-703-7 (1-67) x AB199311 (1-5425)
OY 2 GINGLUPHEMEALALEUGLULEULYSASNLEUPROGLYLUPHESNSERGLYASNGIN 21
Dd ::::::::::::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
729 GAGGAGTTCAATTCTATTCGCATCTGGACACATACAAAGAACCCTTGAACTAGCAAATAA 788
OY 22 PROSEASNAATGGLILYSASNAATGTATGAAGAPILELEUPROPHENIHS----- 38
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
789 GAAGAAAACAGAGAAAAAACAGATACCCACAATCTGCCCAATGATCATTCACAGAGTG 848
OY 39 -----HIS 39
Dd 849 ATTTTGCACCAAGTGATGAATGCCCTGCTGTACTAATTAATCTTCCATCAGAT 908
OY 40 GLTYR-----SERGLYPROBSNGLUATGYTHIR--- 49
Dd 909 GGCTACCAAGAAAAAGAACAAATTCATAGCAGCTCAAGCCCTAAGCAGAGACAGTGAAAT 968
OY 50 ---Phetp 51
Dd 969 GACTTCTGG 977
RESULT 12
AAAT35308
ID AAAT35308 standard; cDNA; 1747 BP.
AC AAAT35308;
XX
```

DT	13-NOV-1996	(first entry)
DE	Rat M1PMP gene cDNA clone derived from rat megakaryocyte cDNA library.	
KM	PMP; protein tyrosine phosphatase; SH2; Src homology region 2;	
KW	chromosome 12p; abnormality; mutation; detection; probe; neoplasia;	
OS	cancer; leukaemia; diagnosis; megakaryocyte regulation; ds.	
XX		
XX	Rattus sp.	
XX		
XX		
XX		
XX	US5536636-A.	
XX		
XX	16-JUL-1996.	
PD		
PF	26-JUN-1991;	91US-0721112.
PR		
PR	28-FEB-1994;	94US-0202389.
PR	26-JUN-1991;	91US-0721112.
PR	31-JAN-1992;	92US-0829141.
PR	01-DEC-1992;	92US-0983926.
PA	(BETH-) BETH ISRAEL HOSPITAL ASSOC.	
PA	(MASI) MASSACHUSETTS INST TECHNOLOGY.	
XX		
PI	Freeman RM, Neel BG, Plutzky J, Rosenberg RD;	
XX		
DR	WPI: 1996-341506/34.	
DR	P-PDB: AAR99314.	
XX		
PT	Detecting 12p chromosomal abnormality associated with neoplastic disease - using SH-PTP1 protein tyrosine phosphatase gene specific probe	
XX		
PS	Example 1; Column 27-32; 63pp; English.	
XX		
CC	AAT35308 is a rat cDNA sequence encoding M1PMP (protein tyrosine phosphatase-1). This sequence was used as a probe to isolate a series of overlapping human cDNA clones defining the human M1PMP equivalent, SH-PTP1 which contains two tandem SH2 domains. A sequence complementary to nucleotides 537-653 of the SH-PTP1 gene clone given as AAT35306 is used as a probe to detect a chromosome 12p13 abnormality associated with neoplastic disease, in partic. CC acute lymphoblastic leukemia. The probe hybridises to a part of the region coding for the two tandem SH2 domains (see AAR99312).	
CC	If the probe will not hybridise DNA of chromosome 12p13 from a patient sample it is indicative of an abnormality, normally associated with CC neoplasia. Fragments of sequences encoding human SH-PTP2 (see AAT35307) may also be used to diagnose a condition or susceptibility to a condition associated with chromosome 12p abnormalities. Alternatively the wild-type SH-PTP1 or SH-PTP2 gene or protein may be used for comparison to sequenced PTP genes taken from a patient, where differences indicate an abnormality. The activity of SH-PTP1 may also be purposely altered by mutation to effect a change in megakaryocyte function and hence platelet production.	
CC		
CC		
XX		
SQ	Sequence 1747 BP; 475 A; 455 C; 485 G; 332 T; 0 other;	
XX		
Alignment Scores:		
Pred. No.:	0.0343	Length: 1747
Score:	88.00	Matches: 19
Percent Similarity:	65.00%	Conservative: 7
Best Local Similarity:	47.50%	Mismatches: 10
Query Match:	23.85%	Indels: 4
DB:	17	Gaps: 2
US-09-935-703-7 (1-67) x AAT35308 (1-1747)		
QY	2 GlnGluPheMetAlaLeu-----GlnLeuLysAnsLeuProGlyGluPheAnsSer 18	

Db 491 GAGGAGTTTGAGAGTGTCTGCAAAAAGACAGAGCTTAAGAACTGCACACGCTCTG---GAA 547
 Oy 19 GtAAsnGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
 548 GGGCAGCGGGCCGAGAACAGACAGACCGCTACAGAACATCTTCCCTTTGACAC 607

RESULT 13

AAT35310
 ID AAT35310 standard; cDNA; 1624 BP.

XX AAT35310;

XX 13-NOV-1996 (first entry)

DE Human SH-PTP1 cDNA clone derived from erythroleukemia cell line.

KM PTP; protein tyrosine phosphatase; SH2; Src homology region 2;

KW chromosome 12p; abnormality; mutation; detection; probe; neoplasia;

KW cancer; leukaemia; diagnosis; megakaryocyte regulation; ds.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS 145..2040

PD 16-JUL-1996.

XX 26-JUN-1991; 91US-0721112.

XX 28-FEB-1994; 94US-0202389.

PR 26-JUN-1991; 91US-0721112.

PR 31-JAN-1992; 92US-0829141.

PR 01-DEC-1992; 92US-0983926.

XX (BETH-) BETH ISRAEL HOSPITAL ASSOC.

PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX Freeman RM, Neel BG, Plutzky J, Rosenberg RD;

DR WPI; 1996-341506/34.

XX P-PSDB; AAR99316.

XX Detecting 12p chromosomal abnormality associated with neoplastic

PT disease - using SH-PTP1 protein tyrosine phosphatase gene specific

PT probe

XX Example 1; Column 43-48; 63pp; English.

XX AAT35310 is a human cDNA sequence encoding a variant SH-PTP1 (protein

CC tyrosine phosphatase-1, with two SH2 domains) obtained from a human

CC erythroleukemia cell line. The variant sequence lacks a sequence

CC of two nucleotides. AG, corresponding to positions 1868 and 1869 of

CC the wild-type sequence (see AAT35306). A fragment complementary to the

CC sequence from nucleotides 537-653 is used as a probe to detect a

CC chromosome 12p13 abnormality associated with neoplastic disease, in

CC partic. acute lymphoblastic leukaemia. The probe hybridises to a part

CC of the region coding for the two tandem SH2 domains (see AAR99112).

CC If the probe will not hybridise DNA of chromosome 12p13 from a patient

CC sample it is indicative of an abnormality, normally associated with

CC neoplasia. Alternatively the wild-type SH-PTP1 or SH-PTP2 gene or

CC protein may be used for comparison to sequenced PTP genes taken from a

CC patient, where differences indicate an abnormality. The activity of

CC SH-PTP1 may also be purposely altered by mutation to effect a change

CC in megakaryocyte function and hence platelet production.

XX Sequence 1624 BP; 381 A; 463 C; 486 G; 294 T; 0 other;

XX Alignment Scores: 0.0592 Length: 1624

Score: 86.00 Matches: 19
 Percent Similarity: 65.00% Conservative: 7
 Best Local Similarity: 47.50% Mismatches: 10
 Query Match: 23.31% Indels: 4
 Db: 17 Gaps: 2

US-09-935-703-7 (1-67) x AAT35310 (1-1624)

Oy 2 GtngluphemeAlaLeu-----GluLeuLysAsnLeuProGluGluPheAsnSer 18

Db 880 GAGGAGTTTGAGAGTGTCTGCAAGAAGCAGAGGTGAAGACTTGCACACGCTCTG---GAA 936

Oy 19 GtAAsnGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38

Db 937 GGGCAGCGGGCCGAGAACAGACAGACCGCTACAGAACATCTTCCCTTTGACAC 996

RESULT 14

AAT26454
 ID AAT26454 standard; cDNA; 1788 BP.

XX AAT26454;

XX 26-MAR-2001 (first entry)

DE SHP-1 DNA.

KM SHP-2; SHP-1; Src Homology-2; protein tyrosine phosphatase; mutant;

KW neoplastic disorder; obesity; angiogenesis; cancer; immune;

KW hematopoietic; allergy; ss.

XX Undetected.

XX US6156551-A.

XX 05-DEC-2000.

XX 05-JUN-1998; 98US-0092443.

XX 05-JUN-1998; 98US-0092443.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

PA (JOSL-) JOSLIN DIABETES CENT.

XX Neel BG, Shoelson S, Pluskey S, O'Reilly AM;

DR WPI; 2001-060166/07.

XX Mutant SH2 domain-containing protein tyrosine phosphatase, useful in in

PT vitro assays to screen for binding partners, inhibitors of tyrosine

PT phosphatase and for treating tyrosine phosphatase-mediated diseases

XX Disclosure; Fig 6; 161pp; English.

XX The present invention relates to an activated SH2 (Src Homology-2)

CC domain containing protein tyrosine phosphatase (SHP-1 or SHP-2)

CC mutant with a mutation in the SH2 domain. Activated mutants of

CC SH2 domain-containing protein tyrosine phosphatases are useful in

CC in vitro assays to screen for binding partners and inhibitors of

CC the phosphatase and in the treatment of PTP-mediated diseases or

CC conditions in a mammal, including neoplastic disorders, obesity and

CC mutants are useful for the treatment of cancer, immunosuppression,

CC immunostimulation, hematopoietic stimulation and anti-allergy

XX treatment.

XX Sequence 1788 BP; 437 A; 490 C; 550 G; 311 T; 0 other;

XX Alignment Scores: 0.0673 Length: 1788

XX Pred. No.: 86.00 Matches: 19

XX Score: 86.00 Conservative: 7

XX Percent Similarity: 65.00% Mismatches: 10

XX Best Local Similarity: 47.50% Indels: 4

XX Query Match: 23.31% Gaps: 2

DB: 22 Gaps: 2
 US-09-935-703-7 (1-67) x AAF26466 (1-1788)
 QY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18
 Db 736 GAGGAGTTTGAGAGTTTGCAGAACGAGGTGGAAGAACTTGCACCGCTCTG---GAA 792
 QY 19 GlyAsnGlnProSerAsnArgGluLysAsnArgTyrArgAspIleuProPheGlnHis 38
 Db 793 GGGCAACGGCCAGAGAACAGGCAAGACCGCTACAGAACATCTCTCCCTTTGACAC 852

QY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18
 Db 736 GAGGAGTTTGAGAGTTTGCAGAACGAGGTGGAAGAACTTGCACCGCTCTG---GAA 792
 QY 19 GlyAsnGlnProSerAsnArgGluLysAsnArgTyrArgAspIleuProPheGlnHis 38
 Db 793 GGGCAACGGCCAGAGAACAGGCAAGACCGCTACAGAACATCTCTCCCTTTGACAC 852

Search completed: June 11, 2003, 12:55:22
 Job time : 215 secs

RESULT 15
 AAF26466
 ID AAF26466 standard; cDNA; 1788 BP.
 XX
 AC AAF26466;
 XX
 DT 26-MAR-2001 (first entry)
 XX
 DE SHP-1 activated double compound mutant D59A/D419A DNA.
 XX
 KW SHP-2: SHP-1: Src Homology-2; protein tyrosine phosphatase; mutant;
 KW neoplastic disorder; obesity; angiogenesis; cancer; immune;
 KW hematopoietic; allergy; ss.
 XX
 OS Unidentified.
 XX
 PN US6156551-A.
 XX
 PD 05-DEC-2000.
 XX
 PF 05-JUN-1998; 98US-0092443.
 XX
 PR 05-JUN-1998; 98US-0092443.
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA (JOSL-) JOSLIN DIABETES CENT.
 PI Neel BG, Shoelson S, Pluskey S, O'Reilly AM;
 PI WPI: 2001-060166/07.
 DR
 XX
 PT Mutant SH2 domain-containing protein tyrosine phosphatase, useful in in
 PT vitro assays to screen for binding partners, inhibitors of tyrosine
 PT phosphatase and for treating tyrosine phosphatase-mediated diseases
 XX
 PS Disclosure; Fig 18; 161pp; English.
 CC
 CC The present invention relates to an activated SH2 (Src Homology-2)
 CC -domain containing protein tyrosine phosphatase (SHP-1 or SHP-2)
 CC mutant with a mutation in the SH2-domain. Activated mutants of
 CC SH2-domain-containing protein tyrosine phosphatases are useful in
 CC in vitro assays to screen for binding partners and inhibitors of
 CC the phosphatase and in the treatment of PTP-mediated diseases or
 CC conditions in a mammal, including neoplastic disorders, obesity and
 CC to inhibit angiogenesis, inhibitors identified using the activated
 CC mutants are useful for the treatment of cancer, immunosuppression,
 CC immunostimulation, hematopoietic stimulation and anti-allergy
 CC treatment.
 CC
 SQ Sequence 1788 BP; 435 A; 492 C; 550 G; 311 T; 0 other;

Alignment Scores:

Pred. No.: 0.0673 Length: 1788
 Score: 86.00 Matches: 19
 Percent Similarity: 65.00% Conservative: 7
 Best Local Similarity: 47.50% Mismatches: 10
 Query Match: 23.31% Indels: 4
 DB: 22 Gaps: 2

US-09-935-703-7 (1-67) x AAF26466 (1-1788)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 11:26:50 ; Search time 11 Seconds

(Without alignments)
252.629 Million cell updates/sec

Title: US-09-935-703-7

Sequence: 1 MOEFMALELNKLPEFNSGN.....TFWHSNEGAVSILRRYCA 67

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88.5	24.0	699	1	PTPE_MOUSE
2	88.5	24.0	700	1	PTPE_MOUSE
3	88	23.8	595	1	PTNE_MOUSE
4	86	23.3	595	1	PTNE_MOUSE
5	79.5	21.5	802	1	PTNE_MOUSE
6	78.5	21.3	458	1	PTNI_HUMAN
7	77.5	21.0	796	1	PTNE_MOUSE
8	77	20.9	521	1	PTPI_DICDI
9	75	20.3	1631	1	PTPI_DROME
10	73.5	19.9	1123	1	PTNE_MOUSE
11	73	19.8	845	1	PTNE_MOUSE
12	72	19.5	585	1	PTNE_MOUSE
13	72	19.5	593	1	PTNE_MOUSE
14	72	19.5	593	1	PTNE_MOUSE
15	71	19.2	775	1	PTNE_MOUSE
16	71	19.2	780	1	PTNE_MOUSE
17	71	19.2	802	1	PTNE_MOUSE
18	70	19.0	363	1	PTNE_MOUSE
19	70	19.0	363	1	PTNE_MOUSE
20	70	19.0	415	1	PTNE_MOUSE
21	69	18.7	1152	1	PTNE_MOUSE
22	69	18.7	1175	1	PTNE_MOUSE
23	69	18.7	1175	1	PTNE_MOUSE
24	69	18.7	1304	1	PTNE_MOUSE
25	68.5	18.6	989	1	PTNE_MOUSE
26	68	18.4	432	1	PTNE_MOUSE
27	68	18.4	432	1	PTNE_MOUSE
28	68	18.4	435	1	PTNE_MOUSE
29	68	18.4	1174	1	PTNE_MOUSE
30	67	18.2	434	1	PTNE_MOUSE
31	67	18.2	829	1	PTNE_MOUSE
32	66	17.9	668	1	PTNE_MOUSE
33	63.5	17.2	807	1	PTNE_MOUSE

34	63.5	17.2	989	1	PTP3_DICDI
35	63	17.1	414	1	PTP1_PSELE
36	63	17.1	593	1	PTNE_MOUSE
37	63	17.1	1189	1	PTNE_MOUSE
38	63	17.1	2029	1	PTNE_MOUSE
39	62	16.8	377	1	PTP2_DICDI
40	62	16.8	444	1	TIG_MYCPN
41	62	16.8	1255	1	CD45_MOUSE
42	62	16.8	2314	1	PTP2_MOUSE
43	62	16.8	2316	1	PTP2_MOUSE
44	61.5	16.7	794	1	YAFH_ECOLI
45	61.5	16.7	1912	1	PTPD_HUMAN

ALIGNMENTS

RESULT 1
PTPE_MOUSE STANDARD: PRT; 699 AA.
AC P49446; Q62134; Q62444; Q64496;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein-tyrosine phosphatase epsilon precursor (EC 3.1.3.48) (R-PTP-epsilon)
GN PTPE OR PTPE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N;
RX MEDLINE=96064677; PubMed=7592814;
RA Elson A., Leder P.;
RT "Protein-tyrosine phosphatase epsilon. An isoform specifically expressed in mouse mammary tumors initiated by v-Ha-ras or neu.";
RL J. Biol. Chem. 270:26116-26122(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DKA/2;
RX Mukoyama Y.;
RA Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; Tissue=Brain, and Lung;
RA Hou E.W., Li S.L.;
RT Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 224-332 FROM N.A.
RC STRAIN=BALB/c; Tissue=Brain;
RX MEDLINE=93086603; PubMed=1454056;
RA Schepens J., Zeeuwen P., Wieringa B., Hendriks W.;
RT "Identification and typing of members of the protein-tyrosine phosphatase gene family expressed in mouse brain.";
RL Mol. Biol. Rep. 16:241-248(1992).
RN [5]
RP SEQUENCE OF 224-332 FROM N.A.
RC STRAIN=BALB/c; Tissue=Brain;
RX MEDLINE=95134232; PubMed=7832766;
RA Hendriks W., Schepens J., Brugman C., Zeeuwen P., Wieringa B.;
RT "A novel receptor-type protein tyrosine phosphatase with a single catalytic domain is specifically expressed in mouse brain.";
RL Blochem. J. 305:499-504(1995).
RN [6]
RP CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein tyrosine + phosphate.
RN [7]
RP SUBCELLULAR LOCATION: Type I membrane protein.
RN [8]
RP SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
RN [9]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U35368; AAC5281.1; -
CC EMBL; D83484; BAA11927.1; -
CC EMBL; U62387; AAB04553.1; -
CC EMBL; 223052; CAA80587.1; -
CC EMBL; 223053; CAA80588.1; -
CC HSSP; P18052; IYFO.
CC MGI; MGI:97813; Ptpre.
CC InterPro; IPR000387; TYR_phosphatase.
CC InterPro; IPR000242; TYR_Pp.
CC Pfam; PF00102; Y-phosphatase; 2.
CC PRINTS; PR00700; PTPPHPTASE.
CC SMART; SM00194; PTPC; 2.
CC PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
CC PROSITE; PS00056; TYR_PHOSPHATASE_2; 2.
CC PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
CC Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat; Signal.
KW SIGNAL.
FT CHAIN 1 19 POTENTIAL.
FT DOMAIN 20 699 PROTEIN-TYROSINE PHOSPHATASE EPSILON.
FT TRANSMEM 46 45 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 69 699 POTENTIAL.
FT DOMAIN 153 392 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 393 699 PROTEIN-TYROSINE PHOSPHATASE 1.
FT ACT_SITE 334 334 PROTEIN-TYROSINE PHOSPHATASE 2.
FT ACT_SITE 629 629 BY SIMILARITY.
FT CARBOHYD 23 23 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 31 31 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CONFLICT 500 500 G -> A (IN REF. 2).
FT CONFLICT 506 506 G -> V (IN REF. 2).
FT CONFLICT 521 522 IV -> ML (IN REF. 2).
FT CONFLICT 606 606 M -> I (IN REF. 1).
SQ SEQUENCE 699 AA; 80645 MW; 4D04467438017FEB CRC64;

Query Match 24.0%; Score 88.5; DB 1; Length 699;
Best Local Similarity 28.9%; Pred. No. 0.0033;
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

QY 2 QEFMALELKLKLPGEFNSGNQPSNRKRYRDLIPFOH-----H 39
DB 136 EEFNSLPSGHIQGFELANKKEENREKNRPILPNDHCIVILSQVDGIPCSDYINASYID 195
QY 40 GY-----SGPNERTT--FW 51
DB 196 GYKEKNKFIAAGPKQETVNDFW 218

RESULT 2
PTPE_HUMAN STANDARD; PRT; 700 AA.
AC P23469;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein-tyrosine phosphatase epsilon precursor (EC 3.1.3.48) (R-PTP-epi1lon).
GN PTPRE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=91006018; PubMed=2170109;
RA Krueger N.X., Streuli M., Saito H.;
RT "Structural diversity and evolution of human receptor-like protein tyrosine phosphatases.";

EMBO J. 9:3241-3252(1990).
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O - protein tyrosine + phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X54134; CAA38069.1; -
CC PIR; S12053; S12053.
CC HSSP; P18052; IYFO.
CC Genew; HGNC:9669; PTPRE.
CC MIM; 600926; -
CC InterPro; IPR000387; TYR_phosphatase.
CC InterPro; IPR000242; TYR_Pp.
CC Pfam; PF00102; Y-phosphatase; 2.
CC PRINTS; PR00700; PTPPHPTASE.
CC SMART; SM00194; PTPC; 2.
CC PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
CC PROSITE; PS00056; TYR_PHOSPHATASE_2; 2.
CC PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
CC Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat; Signal.
KW SIGNAL.
FT CHAIN 1 19 POTENTIAL.
FT DOMAIN 20 700 PROTEIN-TYROSINE PHOSPHATASE EPSILON.
FT TRANSMEM 47 46 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 70 700 POTENTIAL.
FT DOMAIN 154 393 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 394 700 PROTEIN-TYROSINE PHOSPHATASE 1.
FT ACT_SITE 335 335 PROTEIN-TYROSINE PHOSPHATASE 2.
FT ACT_SITE 630 630 BY SIMILARITY.
FT CARBOHYD 23 23 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 30 30 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 700 AA; 80641 MW; D09BCADCEA5708 CRC64;

Query Match 24.0%; Score 88.5; DB 1; Length 700;
Best Local Similarity 28.9%; Pred. No. 0.0033;
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

QY 2 QEFMALELKLKLPGEFNSGNQPSNRKRYRDLIPFOH-----H 39
DB 137 EEFNSLPSGHIQGFELANKKEENREKNRPILPNDHSRVILSOLDGIPCSDYINASYID 196
QY 40 GY-----SGPNERTT--FW 51
DB 197 GYKEKNKFIAAGPKQETVNDFW 219

RESULT 3
PTN6_MOUSE STANDARD; PRT; 595 AA.
AC P29351; Q63872; Q63873; Q63874; Q9QVA6; Q9QVA7; Q9QVA8; Q9ROV6;
AC Q921G3; Q35128;
DT 01-DEC-1992 (Rel. 24, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein-tyrosine phosphatase, non-receptor type 6 (EC 3.1.3.48) (Protein-tyrosine phosphatase 1C) (PTP-1C) (hematopoietic cell de protein-tyrosine phosphatase) (70Z-SHP) (SH-PTP1).
GN PTPN6 OR PTP1C OR HCP OR HCPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

CC STRAIN-RB/2;
RX MEDLINE-92123209; PubMed-1732748;
RT Yi T., Cleveland J.L., Ihle J.N.;
RT "Protein tyrosine phosphatase containing SH2 domains:
RT characterization, preferential expression in hematopoietic cells, and
RT localization to human chromosome 12p12-p13,"
RL Mol. Cell. Biol. 12:836-846(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-92236615; PubMed-1373816;
RA Matthews R.J., Bowne D.B., Flores E., Thomas M.L.;
RT "Characterization of hematopoietic intracellular protein tyrosine
RT phosphatases: description of a phosphatase containing an SH2 domain
RT and another enriched in proline-, glutamic acid-, serine-, and
RT threonine-rich sequences,"
RL Mol. Cell. Biol. 12:2396-2405(1992).
RN [3]
RP SEQUENCE FROM N.A., AND VARIANTS MOTHHEATEN AND VIABLE MOTHHEATEN.
RC STRAIN-C57BL/6J; TISSUE-Bone marrow;
RX MEDLINE-93313972; PubMed-8324828,
RA Schultz L.D., Schweitzer P.A., Kadjan T.V., Yi T., Ihle J.N.,
RA Matthews R.J., Thomas M.L., Belter D.R.;
RT "Mutations at the murine motheaten locus are within the hematopoietic
RT cell protein-tyrosine phosphatase (Hcph) gene,"
RL Cell 73:1445-1454(1993).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).
RX MEDLINE-98112780; PubMed-9445485;
RA Anseri-Lari M.A., Oeltjen J.C., Schwartz S., Zhang Z., Munzy D.M.,
RA Lu J., Gorrell J.H., Chinault A.C., Belmont J.W., Miller W.,
RA Gibbs R.A.;
RT "Comparative sequence analysis of a gene-rich cluster at human
RT chromosome 12p13 and its syntenic region in mouse chromosome 6,"
RL Genome Res. 8:29-40(1998).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND SUBUNIT.
RC STRAIN-C3H; TISSUE-Adrenal gland;
RX MEDLINE-93048302; PubMed-10419485;
RA Martin A., Tsui H.W., Shulman J., Izenman D., Tsui F.W.;
RT "Murine SH-1 splice variants with altered Src homology 2 (SH2)
RT domains. Implications for the SH2-mediated intramolecular regulation
RT of SHP-1,"
RL J. Biol. Chem. 274:21725-21734(1999).
RN [6]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 54-68; 128-135; 137-151; 242-252; 278-285; 293-308 AND
RX 373-382, AND PHOSPHORYLATION.
RP MEDLINE-93054686; PubMed-1385421;
RA Yeung Y.-G., Berg K.L., Plaxley F.J., Angeletti R.H., Stanley E.R.;
RT "Protein tyrosine phosphatase-Jc is rapidly phosphorylated in tyrosine
RT in macrophages in response to colony stimulating factor-1,"
RL J. Biol. Chem. 267:23447-23450(1992).
CC -1- FUNCTION: PLAYS A KEY ROLE IN HEMATOPOIESIS. THIS PTPASE ACTIVITY
CC MAY DIRECTLY LINK GROWTH FACTOR RECEPTORS AND OTHER SIGNALING
CC PROTEINS THROUGH PROTEIN-TYROSINE PHOSPHORYLATION. THE SH2 REGIONS MAY
CC MAY INTERACT WITH OTHER CELLULAR COMPONENTS TO MODULATE ITS OWN
CC PHOSPHATASE ACTIVITY AGAINST INTERACTING SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O -> protein
CC tyrosine + phosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SUBUNIT: Monomer.
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN HEMATOPOIETIC
CC CELLS.
CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; 1 (shown here), 2 and 3; are
CC produced by alternative splicing.
CC -1- PTM: Phosphorylated on tyrosine residues.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 SH2 DOMAINS.
CC -----

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between The Swiss Institute of Bioinformatics and the EMBL outstation -
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL; M68902; AAA37796.1; -.
DR EMBL; M90389; AAA40007.1; -.
DR EMBL; S63763; -. NOT_ANNOTATED_CDS.
DR EMBL; S63764; -. NOT_ANNOTATED_CDS.
DR EMBL; S63803; -. NOT_ANNOTATED_CDS.
DR EMBL; AC002397; AAC36009.1; -.
DR EMBL; AC002397; AAC36008.1; -.
DR EMBL; U65955; AAD00152.1; JOINED.
DR EMBL; U65952; AAD00152.1; JOINED.
DR EMBL; U65953; AAD00152.1; JOINED.
DR EMBL; U65954; AAD00152.1; JOINED.
DR EMBL; U65955; AAD00151.1; JOINED.
DR EMBL; U65951; AAD00151.1; JOINED.
DR EMBL; U65952; AAD00151.1; JOINED.
DR EMBL; U65953; AAD00151.1; JOINED.
DR EMBL; U65954; AAD00151.1; JOINED.
DR EMBL; BC012660; AAH12660.1; -.
DR PIR; A44380; A44390.
DR HSSP; P29350; IGW2.
DR MGD; MG1:9605; Hcph.
DR InterPro; IPR000980; SH2.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000342; TYR_PP.
DR Pfam; PF00017; SH2; 2.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PTYPHPPTASE.
DR PRINTS; PR00401; SH2DOMAIN.
DR ProDom; PD000093; SH2; 2.
DR SMART; SM00194; PTPc; 1.
DR SMART; SM00252; SH2; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
DR PROSITE; PS50001; SH2; 2.
DR Hydrolase; SH2 domain; Repeat: Phosphorylation; Alternative splicing.
FT DOMAIN 4 100 SH2 1.
FT FT 110 213 SH2 2.
FT FT 269 514 PROTEIN-TYROSINE PHOSPHATASE.
FT FT ACT_SITE 453 453 BY SIMILARITY.
FT FT VARSELIC 1 39 MYR -> MSTRG (IN ISOFORM 2).
FT FT VARSLIC 40 44 MISSING (IN ISOFORM 3).
FT FT VARIANT 77 99 SVSVR -> MSTRG (IN ISOFORM 3).
FT FT VARIANT 100 595 EYTYDQGLDQRDGIITLKTP -> VVRPIHRAAGVTA
FT FT CONFLICT 240 240 ACGRGALD (IN MOTHEATEN (ME)).
FT FT CONFLICT 572 572 MISSING (IN MOTHEATEN (ME)).
FT FT CONFLICT 586 586 K -> R (IN REF. 1 AND 3).
FT FT CONFLICT 586 586 E -> Q (TN REF. 1 AND 3).
FT FT CONFLICT 586 586 E -> D (TN REF. 6).
SQ SEQUENCE 595 AA; 67559 MW; CFLI7300D0J2638D2 CRC64;

Query Match 23.8%; Score 88; DB 1; Length 595;
Best Local Similarity 47.5%; Pred. No. 0.0031;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2

Db 2 OEPMAL--ELKNLPGEFNSGNQPSNREKRRYDLIPFOH 38
      :||:| | | | | | | | | | | | | | | | |
246 EEFEESLQKEVKNLHORL-EGQPENKSKRKRYKNLIPFDH 284

RESULT 4
ID PTNG_HUMAN STANDARD; PRT; 595 AA.
AC P29350;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
```

DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE Protein-tyrosine phosphatase; non-receptor type 6 (EC 3.1.3.48)
 DE (Protein-tyrosine phosphatase 1C) (PTP-1C) (Hematopoietic cell
 DE SHP-1)
 GN PTPN6 OR PTP1C OR HCP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RX MEDLINE-92123209; PubMed-1732748;
 RA Yi T., Cleveland J.L., Ihle J.N.;
 RT "Protein tyrosine phosphatase containing SH2 domains:
 RT characterization, preferential expression in hematopoietic cells, and
 RT localization to human chromosome 12p12-p13.";
 RL Mol. Cell. Biol. 12:836-846(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RX TISSUE-Breast;
 RC MEDLINE-91343005; PubMed-1652101;
 RA Shen S.H., Bastien L., Posner B.I., Chretien P.;
 RT "A protein-tyrosine phosphatase with sequence similarity to the SH2
 RT domain of the protein-tyrosine kinases.";
 RL Nature 352:736-739(1991).
 RN [3]
 RP REVISIONS.
 RA Shen S.H., Bastien L., Posner B.I., Chretien P.;
 RL Nature 353:868-868(1991).
 RN [4]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RX MEDLINE-92141214; PubMed-1736296;
 RA Plutsky J., Neel B.G., Rosenberg R.D.;
 RT "Isolation of a src homology 2-containing tyrosine phosphatase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:1123-1127(1992).
 RN [5]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RX MEDLINE-95394454; PubMed-765165;
 RA Banville D., Stocco R., Shen S.H.;
 RT "Human protein tyrosine phosphatase 1C (PTPN6) gene structure:
 RT alternate promoter usage and exon skipping generating multiple
 RT transcripts.";
 RL Genomics 27:165-173(1995).
 RN [6]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RX MEDLINE-96303695; PubMed-8723724;
 RA Ansari-Lari M.A., Muzny D.M., Lu J., Lu F., Lilley C.E., Spanos S.,
 RT "A gene-rich cluster between the CD4 and triosephosphate isomerase
 RT genes at human chromosome 12p13.";
 RL Genome Res. 6:314-326(1996).
 RN [7]
 RP PHOSPHORYLATION.
 RX MEDLINE-95300784; PubMed-7781604;
 RA Li R.Y., Gault F., Ragab-Thomas J.M.F., Chap H.;
 RT "Tyrosine phosphorylation of an SH2-containing protein tyrosine
 RT phosphatase is coupled to platelet thrombin receptor via a pertussis
 RT toxin-sensitive heterotrimeric G-protein.";
 RL EMBO J. 14:2519-2526(1995).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 248-399.
 RX MEDLINE-98447672; PubMed-9774441;
 RA Yang J., Liang X., Niu T., Meng W., Zhao Z., Zhou G.W.;
 RT "Crystal structure of the catalytic domain of protein-tyrosine
 RT phosphatase SHP-1.";
 RL J. Biol. Chem. 273:28199-28207(1998).
 CC -1- FUNCTION: PLAYS A KEY ROLE IN HEMATOPOIESIS. THIS PTASE ACTIVITY
 CC MAY DIRECTLY LINK GROWTH FACTOR RECEPTORS AND OTHER SIGNALING
 CC PROTEINS THROUGH PROTEIN-TYROSINE PHOSPHORYLATION. THE SH2 REGIONS
 CC MAY INTERACT WITH OTHER CELLULAR COMPONENTS TO MODULATE ITS OWN
 CC PHOSPHATASE ACTIVITY AGAINST INTERACTING SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O -> protein

CC tyrosine + phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE)
 CC AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN HEMATOPOIETIC
 CC CELLS.
 CC -1- PTM: PHOSPHORYLATED ON SERINE AND TYROSINE.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 SH2 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC -----
 DR EMBL: M74903; AAA35963.1; -;
 DR EMBL: X62055; CAA43982.1; -;
 DR EMBL: M77273; AAA36610.1; -;
 DR EMBL: U15528; AAA82880.1; -;
 DR EMBL: U15536; AAA82880.1; JOINED.
 DR EMBL: U15535; AAA82880.1; JOINED.
 DR EMBL: U15534; AAA82880.1; JOINED.
 DR EMBL: U15533; AAA82880.1; JOINED.
 DR EMBL: U15532; AAA82880.1; JOINED.
 DR EMBL: U15531; AAA82880.1; JOINED.
 DR EMBL: U15529; AAA82880.1; JOINED.
 DR EMBL: U15528; AAA82880.1; JOINED.
 DR EMBL: U15527; AAA82879.1; JOINED.
 DR EMBL: U15535; AAA82879.1; JOINED.
 DR EMBL: U15534; AAA82879.1; JOINED.
 DR EMBL: U15533; AAA82879.1; JOINED.
 DR EMBL: U15531; AAA82879.1; JOINED.
 DR EMBL: U15530; AAA82879.1; JOINED.
 DR EMBL: U15529; AAA82879.1; JOINED.
 DR EMBL: U47924; AAB51323.1; -;
 DR EMBL: U47924; AAB51322.1; -;
 DR PIR: S20825; S20825.
 DR PIR: B42031; B42031.
 DR PDB: 1GM2; 22-AUG-99.
 DR PDB: 1EPR; 07-MAR-01.
 DR Genew; HGNC:9658; PTPN6.
 DR MIM: 176883; -;
 DR InterPro; IPR000980; SH2.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR InterPro; IPR000242; TYR_pp.
 DR Pfam; PF00017; SH2; 2.
 DR Pfam; PF00102; Y_phosphatase; 1.
 DR PRINTS; PR00700; PTPYPPHPTASE.
 DR PRINTS; PR00401; SH2DOMAIN.
 DR Prodom; PD000093; SH2; 2.
 DR SMART; SM00194; PTPc; 1.
 DR SMART; SM00252; SH2; 2.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
 DR PROSITE; PS50001; SH2; 2.
 KW Hydroxylase; SH2 domain; Repeat; Phosphorylation; Alternative splicing;
 KW 3D-structure.
 FT DOMAIN 4 100 SH2 1.
 FT DOMAIN 110 213 SH2 2.
 FT ACT_SITE 269 514 PROTEIN-TYROSINE PHOSPHATASE.
 FT ACT_SITE 453 453
 FT VARSPIC 1 39 MISSING (IN SHORT ISOFORM).
 FT VARSPIC 40 44 SLVR -> MLSRG (IN SHORT ISOFORM).
 FT CONFLICT 1 3 MVR -> MLSRG (IN REF. 2).
 FT CONFLICT 6 6 H -> L (IN REF. 5).

```

OY      2  OEPMAL---ELKKNLPGENGSCNOPSREKKNRYDRIIPQH 38
DB      246  EEFEESLQKEVKNLHQR-EGQRPENKGNKRNKILPDDH 284

RESULT 5
PTXA_HUMAN          STANDARD:          PRT:          802 AA.
AC      P18433: 014513:
DT      01-NOV-1980 (Rel. 16, Created)
DT      01-NOV-1990 (Rel. 16, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Protein-tyrosine phosphatase alpha precursor (EC 3.1.3.48) (R-PMP-
DE      alpha).
GN      PTPA.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90349565; PubMed=2166945;
RA      Sap J., D'Eustachio P., Glyol D., Schlessinger J.;
RT      "Cloning and expression of a widely expressed receptor tyrosine
RT      phosphatase.";
RL      Proc. Natl. Acad. Sci. U.S.A. 87:6112-6116(1990).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90384936; PubMed=2169617;
RA      Kaplan R., Morse B., Huebner K., Croce C., Hawk R., Ravera M.,
RA      Ricca G., Jaye M., Schlessinger J.;
RT      "Cloning of three human tyrosine phosphatases reveals a multigene
RT      family of receptor-linked protein-tyrosine-phosphatases expressed in
RT      brain.";
RL      Proc. Natl. Acad. Sci. U.S.A. 87:7000-7004(1990).
RN      [3]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=91006018; PubMed=2170109;
RA      Krueger N.X., Streuli M., Saito H.;
RT      "Structural diversity and evolution of human receptor-like protein
RT      tyrosine phosphatases.";
RL      EMBO J. 9:3241-3252(1990).
RN      [4]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=91088320; PubMed=2175690;
RA      Ohagi S., Nishi M., Steiner D.F.;
RT      "Sequence of a cDNA encoding human LRP (leukocyte common antigen-
RT      related peptide).";
RL      Nucleic Acids Res. 18:7159-7159(1990).
RN      [5]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=91032191; PubMed=2172030;
RA      Jirik F.R., Janzen N.M., Melhado I.G., Harder K.W.;
RT      "Cloning and chromosomal assignment of a widely expressed human
RT      receptor-like protein-tyrosine phosphatase.";
RL      FEBS Lett. 273:239-242(1990).
RN      [6]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=21638749; PubMed=11780052;
RA      Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA      Jones M., Stavrides G., Almeida A.K., Bagguley C.L.,
RA      Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
RA      Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA      Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,

```

FT CONFLICT 122 122 E -> P (IN REF. 5).
 FT CONFLICT 138 138 S -> SCNSKSKDR (IN REF. 2).
 FT CONFLICT 179 187 MISSING (IN REF. 2, 3, 4, 5 AND 6).
 FT CONFLICT 289 289 G -> E (IN REF. 4).
 FT CONFLICT 367 367 V -> A (IN REF. 4).
 FT CONFLICT 453 493 F -> S (IN REF. 4).
 FT CONFLICT 786 786 K -> E (IN REF. 4).
 SQ SEQUENCE 802 AA; 90599 MW; 8E964CB56B5BE32 CRC64;
 Query Match
 Best Local Similarity 21.5%; Score 79.5; DB 1; Length 802;
 Matches 22; Conservative 9; Mismatches 19; Indels 33; Gaps 3;
 OY 2 QEFALLETNLPGEFNSGNOPSREKRRYRDILPFH-----H 39
 DB 243 EEFNLPACPIATCEASKEERKRRYVILPYDHSRVHLPEGVPSDYINASFIN 302
 OY 40 GY-----SGPNERT--FM 51
 DB 303 GYQEKNFIAAGPKKEFTVNDPM 325
 RESULT 6
 PTNI_HUMAN STANDARD; PRT; 458 AA.
 AC 099952;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Protein-tyrosine phosphatase, non-receptor type 18 (EC 3.1.3.48)
 GN (Brain-derived phosphatase).
 GN PTPN18 OR BDPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCBL_TaxID-9606;
 RN [1]
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND CHARACTERIZATION.
 RC TISSUE-BRAIN;
 RA MEDLINE-97108674; PubMed-8950995;
 RA Kim Y.W., Wang H.Y., Sures I., Lammers R., Martell K.J., Ullrich A.;
 RT Characterization of the PEST family protein tyrosine phosphatase
 RT BDPL.
 RT Oncogene 13:2275-2279(1996).
 CC -1- FUNCTION: Differentially dephosphorylate autophosphorylated
 CC tyrosine kinases which are known to be overexpressed in tumor
 CC tissues.
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
 CC tyrosine + phosphate.
 CC -1- TISSUE SPECIFICITY: Expressed in brain, colon and several tumor-
 CC derived cell lines.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL; X79568; CA56105.1; -;
 DR HSSP; 006124; 2SHP.
 DR Genew; HGNC:9649; PTPN18.
 DR MIM; 606587;
 DR InterPro; IPR000387; TYR_PHSPTASE.
 DR InterPro; IPR000242; TYR_PP.
 DR Pfam; PF00102; Y_PHSPTASE; 1.
 DR PRINTS; PRO0700; PRTYPHPRASE.
 DR SMART; SM00194; PTPC; 1.
 DR PROSITE; PS00383; TYR_PHSPTASE_1; 1.
 DR PROSITE; PS00356; TYR_PHSPTASE_2; 1.
 DR PROSITE; PS00356; TYR_PHSPTASE_2; 1.

DR PROSITE; PS00355; TYR_PHSPTASE_PTP; 1.
 KW Hydrolase.
 FT DOMAIN 26 291 PROTEIN-TYROSINE PHOSPHATASE.
 FT ACT_SITE 229 229 BY SIMILARITY.
 SQ SEQUENCE 458 AA; 50384 MW; 46BCA1E17C2C78B1 CRC64;
 Query Match
 Best Local Similarity 21.3%; Score 78.5; DB 1; Length 458;
 Matches 26; Conservative 10; Mismatches 19; Indels 51; Gaps 4;
 OY 12 LPGEF-----SGNOPSREKRRYRDILPF-----OHGYS- 42
 DB 26 LAGEFSDIQACSAAMKADGCVSTVAGSRPEYKRRYKDYLPDQTRVILSLQEGHSD 85
 OY 43 -----GPNERT--TFHGSVEGAVSLLRRC 66
 DB 86 YINGNFIRGVDSLAYIATGCPPLPHTLDFRLWVEFGVILMAC 131
 RESULT 7
 PTNR_RAT STANDARD; PRT; 796 AA.
 AC 003348;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DE Protein-tyrosine phosphatase alpha precursor (EC 3.1.3.48) (R-PTP-
 DE alpha).
 GN PTPRA OR LRP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NC NCBL_TaxID-10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RA MEDLINE-93036682; PubMed-1417854;
 RA Moriyma T., Fujiwara Y., Imai E., Takenaka M., Kawanishi S.,
 RA Inoue T., Noguchi T., Tanaka T., Kamada T., Ueda N.;
 RT cDNA cloning of rat LRP, a receptor like protein tyrosine
 RT phosphatase, and evidence for its gene regulation in cultured rat
 RT mesangial cells.
 RT Biochem. Biophys. Res. Commun. 188:34-39(1992).
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
 CC tyrosine + phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL; L01702; AAA41983.1; -;
 DR HSSP; P18052; IYFO.
 DR InterPro; IPR000387; TYR_PHSPTASE.
 DR InterPro; IPR000242; TYR_PP.
 DR Pfam; PF00102; Y_PHSPTASE; 2.
 DR PRINTS; PRO0700; PRTYPHPRASE.
 DR SMART; SM00194; PTPC; 2.
 DR PROSITE; PS00383; TYR_PHSPTASE_1; 2.
 DR PROSITE; PS00356; TYR_PHSPTASE_2; 2.
 DR PROSITE; PS00355; TYR_PHSPTASE_PTP; 2.
 KW Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat;
 KW Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 796 PROTEIN-TYROSINE PHOSPHATASE ALPHA.
 FT DOMAIN 20 145 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 146 169 POTENTIAL.
 FT DOMAIN 170 796 CYTOPLASMIC (POTENTIAL).

```

FT DOMAIN 224 494 PROTEIN-TYROSINE PHOSPHATASE 1.
FT ACT_SITE 435 796 PROTEIN-TYROSINE PHOSPHATASE 2.
FT ACT_SITE 436 796 BY SIMILARITY.
FT CARBOHYD 726 726 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 20 20 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 86 86 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 796 AA; 90260 MW; 4793796191056920 CRC64;

Query Match 21.0%; Score 77.5; DB 1; Length 796;
Best Local Similarity 25.3%; Pred. No. 0.088;
Matches 21; Conservative 10; Mismatches 19; Indels 33; Gaps 3;

QY 2 QEFMALELNLPGEFNSGNQPSNREKRRYRDILPFOH-----H 39
DB 237 EEFNLEPACPIATCEAAKSEKREKRYVLLPYDHSRVHLTPVEGVPSDYINASEFIN 296
QY 40 GT-----SGNERT--FW 51
DB 297 GYQENKRFIAOGPKETVDFW 319

RESULT 8
PTPL_DICDI STANDARD; PRT; 521 AA.
ID PTPL_DICDI
AC P34137;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Protein-tyrosine phosphatase 1 (EC 3.1.3.48) (Protein-tyrosine-
  phosphate phosphodiesterase 1).
GN PTPL OR PTPL.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellidae; Dictyostellium.
OX NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93046662; PubMed-1423620;
RA Howard P.K., Sefton B.M., Firtel R.A.;
RT Analysis of a spatially regulated phosphotyrosine phosphatase
  identifies tyrosine phosphorylation as a key regulatory pathway in
  Dictyostelium.
RT Cell 71:637-647(1992).
RL -1- FUNCTION: MAY HAVE A ROLE IN GROWTH AND IN THE EARLY STAGES OF
  DEVELOPMENT. AFFECTS THE TIMING OF DEVELOPMENT.
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
  tyrosine + phosphate.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. MIGHT BE TETHERED TO THE
  PLASMA MEMBRANE.
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN ANTERIOR-LIKE CELLS
  AND TO A LESSER DEGREE IN PRESTALK CELLS.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT A VERY LOW LEVEL IN VEGETATIVE
  CELLS, INDUCED BY 4 HRS, MAXIMALLY EXPRESSED AT THE TIGHT
  AGGREGATE STAGE AND THROUGH THE REMAINDER OF DEVELOPMENT.
CC -1- MISCELLANEOUS: THE PTASE DOMAIN IS INTERRUPTED BY A PTASE INSERT
  WHICH SHARES NO HOMOLOGIES WITH OTHER PTASE PROTEINS.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
  TYROSINE PHOSPHATASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).

```

```

DR EMBL: L07125; AAA33241.1; .
DR HSSP: Q06124; 25HP.
DR Dictydb: DD03018; PTPL.
DR InterPro: IPR000340; DS_phosphatase.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; TYR_PP.
DR Pfam: PF00102; Y_phosphatase.1.
DR PRINTS: PR00700; PRTYPHPTASE.
DR SMART: SM00194; PTPC.1.
DR SMART: SM00012; PTPC-DSPC.1.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP.1.
DR Hydrolase.
FT DOMAIN 1 114 SER-RICH.
FT ACT_SITE 310 310 BY SIMILARITY.
FT DOMAIN 327 425 PTASE INSERT (ASN-RICH).
FT DOMAIN 382 400 POLY-ASN.
SQ SEQUENCE 521 AA; 59427 MW; 0F516AEDD75EAB96 CRC64;

Query Match 20.9%; Score 77; DB 1; Length 521;
Best Local Similarity 43.2%; Pred. No. 0.062;
Matches 16; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

QY 2 QEFMALELNLPGEFNSGNQPSNREKRRYRDILPFOH 38
DB 99 EEFLEESVGPSTSEGDKNHTSKNRTNLPVNH 135

RESULT 9
PTPL_DROME STANDARD; PRT; 1631 AA.
ID PTPL_DROME
AC P35992;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein-tyrosine phosphatase 10D precursor (EC 3.1.3.48) (Receptor-
  linked protein-tyrosine phosphatase 10D).
GN PTPL10D.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
  Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RX TISSUE-Embryo;
RC MEDLINE-92034989; PubMed-1657402;
RA Tian S.-S., Tsoulfas P., Zinn K.;
RT Three receptor-linked protein-tyrosine phosphatases are selectively
  expressed on central nervous system axons in the Drosophila embryo.
RT Cell 67:675-685(1991).
RL [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX TISSUE-Embryo;
RC MEDLINE-92034988; PubMed-1657401;
RA Yang X., Seow K.T., Bahrt S.M., Oon S.H., Chia W.;
RT Two Drosophila receptor-like tyrosine phosphatase genes are
  expressed in a subset of developing axons and pioneer neurons in the
  embryonic CNS.
RL Cell 67:661-673(1991).
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
  tyrosine + phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms, a long form (shown here) and a
  short form; are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN A SUBSET OF AXONS AND
  PIONEER NEURONS IN THE EMBRYO.
CC -1- SIMILARITY: CONTAINS 12 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

[illegible]

Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

-1- FUNCTION: SPECIFICALLY BINDS POLY(G) RNA HOMOPOLYMERS IN VITRO.

-1- SUBCELLULAR LOCATION: Nucleus (Cytoplasmic Fractions).

[illegible]

```

RESULT 12
PTNB_MOUSE
ID PTNB_MOUSE
AC P35235; STANDARD; PRT; 585 AA.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein-tyrosine phosphatase, non-receptor type 11 (EC 3.1.3.48)
GN PRP11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93206094; PubMed=8096088;
RX Feng G.-S., Hui C.-C., Pawson T.,
RX "SH2-containing phosphotyrosine phosphatase as a target of protein-
RX tyrosine kinases.",
RX Science 259:1607-1611(1993).
[2]
RP X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS) OF 4-103.
RX MEDLINE=94363243; PubMed=7521735;
RX Lee C.-H., Komlós D., Jacques S., Margolis B., Schlessinger J.,
RX Shelson S.E., Kurlyan J.;
RX "Crystal structures of peptide complexes of the amino-terminal SH2
RX domain of the SH2 tyrosine phosphatase.",
RX Structure 2:423-438(1994).
RT -1- FUNCTION: THIS PRPASE ACTIVITY MAY DIRECTLY LINK GROWTH FACTOR
RT RECEPTORS AND OTHER SIGNALING PROTEINS THROUGH PROTEIN-TYROSINE
RT PHOSPHORYLATION. THE SH2 REGIONS MAY INTERACT WITH OTHER CELLULAR
RT COMPONENTS TO MODULATE ITS OWN PHOSPHATASE ACTIVITY AGAINST
RT INTERACTING SUBSTRATES.
RT -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
RT tyrosine + phosphate.
RT -1- SUBCELLULAR LOCATION: Cytoplasmic.
RT -1- PTM: PHOSPHORYLATED BY TYROSINE-PROTEIN KINASES.
RT -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
RT TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.
RT -1- SIMILARITY: CONTRAINS 2 SH2 DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L08663; -, NOT_ANNOTATED_CDS

```

DR PIR: A46209; A46209.
 DR PDB: 1A5A; 31-AUG-94.
 DR PDB: 1AYB; 31-AUG-94.
 DR PDB: 1AYC; 31-AUG-94.
 DR PDB: 1AYD; 31-AUG-94.
 DR MGD: MGI:99511; Pcpn11.
 DR InterPro: IPR000980; SH2.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR InterPro: IPR000242; Tyr_PP.
 DR Pfam: PF00017; SH2; 2.
 DR Pfam: PF00102; X-phosphatase; 1.
 DR PRINTS: PRO0700; PRTPHPTASE.
 DR PRINTS: PRO0401; SH2DOMAIN.
 DR PRODOM: PD000093; SH2; 2.
 DR SMART: SM00194; PTPC; 1.
 DR SMART: SM00252; SH2; 2.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS00055; TYR_PHOSPHATASE_PTP; 1.
 DR PROSITE: PS50001; SH2; 2.
 DR Hydrolase: SH2 domain; Repeat; Phosphorylation; 3D-structure.
 FT DOMAIN 6 102 SH2 1.
 FT DOMAIN 112 216 SH2 2.
 FT DOMAIN 276 521 PROTEIN-TYROSINE PHOSPHATASE.
 FT ACT_SITE 463 463 BY SIMILARITY.
 FT SEQUENCE 585 AA; 66816 MW; 6CE554F923B8F72A CRC64;

Query Match 19.5% Score 72; DB 1; Length 585;
 Best Local Similarity 38.1%; Pred. No. 0.3;
 Matches 16; Conservative 7; Mismatches 13; Indels 6; Gaps 2;

OY 19 GNPSNREKNRYRDLIPQH-----HGSGPNERTFTHGNS 55
 Db 268 GORBNKNNKRYKNILPFQTRVRLHD-GDPNEPVSDYINAN 308

RESULT 13
 PTNB_HUMAN STANDARD; PRT; 593 AA.
 AC 006124;
 DT 01-FEB-1994 (Rel. 28; Created)
 DT 01-FEB-1994 (Rel. 28; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Protein-tyrosine phosphatase, non-receptor type 11 (EC 3.1.3.48)
 DE (Protein-tyrosine phosphatase 2C) (PTP-2C) (PTP-1D) (SH-PTP3) (SH-PTP2) (SH-2).
 GN PTPN11 OR PTP2C OR SHPTP2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Umbilical cord;
 RX MEDLINE-93211929; PubMed-7681589;
 RA Ahmad S., Banville D.L., Zhao Z., Fischer E.H., Shan S.H.;
 RT "A widely expressed human protein-tyrosine phosphatase containing src
 RT homology 2 domains";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:2197-2201(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-93206095; PubMed-7681217;
 RA Vogel W., Lammers R., Huang J., Ullrich A.;
 RT "Activation of a phosphotyrosine phosphatase by tyrosine
 RT phosphorylation";
 RL Science 259:1611-1614(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-T-cell;
 RX MEDLINE-93106179; PubMed-1281790;
 RA Adachi M., Sekiya M., Miyachi T., Matsuno K., Hinoda Y., Imai K.,
 RA Yachi A.;
 RT "Molecular cloning of a novel protein-tyrosine phosphatase SH-PTP3

RT with sequence similarity to the src-homology region 2.";
 RL FEBS Lett. 314:335-339(1992).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-94029983; PubMed-8216283;
 RA Bastien L., Ramchandran C., Liu S., Adam M.;
 RT "Cloning, expression and mutational analysis of SH-PTP2, human
 RT protein-tyrosine phosphatase";
 RL Biochem. Biophys. Res. Commun. 196:124-133(1993).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93087502; PubMed-1280823;
 RA Freeman R.M., Jr., Plutsky J., Neel B.G.;
 RT "Identification of a human src homology 2-containing protein-tyrosine-
 RT phosphatase: a putative homolog of Drosophila corkscrew";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:11239-11243(1992).
 RN [6]
 RP PHOSPHORYLATION BY PDGFR.
 RX MEDLINE-94316690; PubMed-8041791;
 RA Bennett A.M., Tang T.L., Sugimoto S., Walsh C.T., Neel B.G.;
 RT "Protein-tyrosine phosphatase SHPTP2 couples platelet-derived growth
 RT factor receptor beta to Ras";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7335-7339(1994).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 1-526.
 RX MEDLINE-98150850; PubMed-9491886;
 RA Hof P., Plutsky S., Dhe-Paganon S., Eck M.J., Shoelson S.E.;
 RT "Crystal structure of the tyrosine phosphatase SHP-2";
 RL Cell 92:441-450(1998).
 CC -1- FUNCTION: THIS PTPASE ACTIVITY MAY DIRECTLY LINK GROWTH FACTOR
 CC RECEPTORS AND OTHER SIGNALING PROTEINS THROUGH PROTEIN-TYROSINE
 CC PHOSPHORYLATION. THE SH2 REGIONS MAY INTERACT WITH OTHER CELLULAR
 CC COMPONENTS TO MODULATE ITS OWN PHOSPHATASE ACTIVITY AGAINST
 CC INTERACTING SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
 CC tyrosine + phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN HUMAN TISSUES.
 CC -1- PARTICULARLY ABUNDANT IN HEART, BRAIN, AND SKELETAL MUSCLE.
 CC -1- PTM: PHOSPHORYLATION OF TYR RESIDUES AT THE C TERMINUS BY
 CC PLATELET-DERIVED GROWTH FACTOR CREATES A BINDING SITE FOR THE SH2
 CC DOMAIN OF GRB2.
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 SH2 DOMAINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sib.ch).
 CC -----
 CC EMBL; L08807; -; NOT ANNOTATED_CDS.
 CC EMBL; X70766; CA50045.1; -;
 CC EMBL; D13540; BAA02740.2; -;
 CC EMBL; L07527; AAA17022.1; -;
 CC EMBL; L03535; AAA36611.1; -;
 CC DR PIR; J08055; J08055.
 CC DR PDB; 2SHP; 16-FEB-99.
 CC DR Genew; HGNC:9644; PTPN11.
 CC DR MIM; 176876; -;
 CC DR InterPro: IPR000980; SH2.
 CC DR InterPro: IPR000387; TYR_phosphatase.
 CC DR InterPro: IPR000242; Tyr_PP.
 CC DR Pfam; PF00017; SH2; 2.
 CC DR Pfam; PF00102; Y_phosphatase; 1.
 CC DR PRINTS; PRO0700; PRTPHPTASE.
 CC DR PRINTS; PRO0401; SH2DOMAIN.
 CC DR PRODOM; PD000093; SH2; 2.
 CC DR SMART; SM00194; PTPC; 1.
 CC DR SMART; SM00252; SH2; 2.

DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
 DR PROSITE; PS50001; SH2; 2.
 KW Hydrolyase; SH2 domain; Repeat; Phosphorylation; 3D-structure.
 FT DOMAIN 6 102 SH2 1.
 FT DOMAIN 112 216 SH2 2.
 FT ACT_SITE 276 517 PROTEIN-TYROSINE PHOSPHATASE.
 FT MOD_RES 542 542 PHOSPHORYLATION (BY PDGFR).
 FT MOD_RES 580 580 PHOSPHORYLATION (BY PDGFR).
 FT CONFLICT 535 535 S -> R (IN REF. 3).
 FT CONFLICT 548 548 S -> P (IN REF. 3).
 SQ SEQUENCE 593 AA; 68011 MW; 9CDBEF5A56CCB45 CRC64;
 Query Match 19.5%; Score 72; DB 1; Length 593;
 Best Local Similarity 38.1%; Pred. No. 0.3;
 Matches 16; Conservative 7; Mismatches 13; Indels 6; Gaps 2;
 QY 19 GNQPSNREKNRYDILPFQH-----HGSGPNERTEFWHGSN 55
 DB 268 GQROENKNNKRYKNIILPFDHTRVYLHD-GDPNEPVSDYINAN 308
 ID PTNB_RAT STANDARD; PRT; 593 AA.
 AC P41499; Q62626;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein-tyrosine phosphatase, non-receptor type 11 (EC 3.1.3.48)
 DE (Protein-tyrosine phosphatase SYP).
 GN PTPN11.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Sprague-Dawley;
 RX MEDLINE-94324984; PubMed-8048963;
 RA Ding W., Zhang W.R., Sullivan K., Hashimoto N., Goldstein B.J.;
 RT "Identification of protein-tyrosine phosphatases prevalent in
 RT adipocytes by molecular cloning.";
 RL Biochem. Biophys. Res. Commun. 202:902-907(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Sprague-Dawley;
 RX MEDLINE-94216346; PubMed-7512964;
 RA Mel L., Dornier C.A., Huganir R.L.;
 RT "RNA splicing regulates the activity of a SH2 domain-containing
 RT protein tyrosine phosphatase.";
 RL J. Biol. Chem. 269:12254-12262(1994).
 RN [3]
 RP PARTIAL SEQUENCE.
 RX MEDLINE-93011127; PubMed-1382983;
 RA Hirata A., Munakata H., Hata K., Suzuki Y., Tsukui S.;
 RT "Purification and characterization of a rat liver protein-tyrosine
 RT phosphatase with sequence similarity to src-homology region 2.";
 RL Eur. J. Biochem. 209:195-206(1992).
 RN [4]
 RP FUNCTION: THIS PTPASE ACTIVITY MAY DIRECTLY LINK GROWTH FACTOR
 RP RECEPTORS AND OTHER SIGNALING PROTEINS THROUGH PROTEIN-TYROSINE
 RP PHOSPHORYLATION. THE SH2 REGIONS MAY INTERACT WITH OTHER CELLULAR
 RP COMPONENTS TO MODULATE ITS OWN PHOSPHATASE ACTIVITY AGAINST
 RP INTERACTING SUBSTRATES.
 CC CATALYTIC ACTIVITY: Protein tyrosine phosphatase
 CC tyrosine + phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- PTM: PHOSPHORYLATED BY TYROSINE-PROTEIN KINASES (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASSES OF THE PROTEIN-
 CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 SH2 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U09307; AAA20543.1; -
 DR EMBL; U05963; AAA19133.1; -
 DR PIR; S29281; S29281.
 DR HSP; P35235; IATN.
 DR InterPro; IPR000980; SH2.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR InterPro; IPR00242; Tyr_Pp.
 DR Pfam; PF00017; SH2; 2.
 DR Pfam; PF00102; Y_phosphatase; 1.
 DR PRINTS; PR00700; PRTYPHRTASE.
 DR PRINTS; PR00401; SH2DOMAIN.
 DR PRODOM; PD000093; SH2; 1.
 DR SMART; SM00194; PTPC; 1.
 DR SMART; SM00252; SH2; 2.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
 DR PROSITE; PS50001; SH2; 2.
 KW Hydrolyase; SH2 domain; Repeat; Phosphorylation.
 FT DOMAIN 6 102 SH2 1.
 FT DOMAIN 112 216 SH2 2.
 FT ACT_SITE 276 517 PROTEIN-TYROSINE PHOSPHATASE.
 FT CONFLICT 459 459 BY SIMILARITY.
 FT CONFLICT 75 75 A -> P (IN REF. 1).
 FT CONFLICT 407 407 G -> GOALL (IN REF. 2).
 FT CONFLICT 547 547 Y -> S (IN REF. 2).
 SQ SEQUENCE 593 AA; 68033 MW; 3329F10F060AF46 CRC64;
 Query Match 19.5%; Score 72; DB 1; Length 593;
 Best Local Similarity 38.1%; Pred. No. 0.3;
 Matches 16; Conservative 7; Mismatches 13; Indels 6; Gaps 2;
 QY 19 GNQPSNREKNRYDILPFQH-----HGSGPNERTEFWHGSN 55
 DB 268 GQROENKNNKRYKNIILPFDHTRVYLHD-GDPNEPVSDYINAN 308
 ID PTNC_MOUSE STANDARD; PRT; 775 AA.
 AC P35831;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein-tyrosine phosphatase, non-receptor type 12 (EC 3.1.3.48)
 DE (Protein-tyrosine phosphatase P19) (P19-PTP) (MPTP-PTP).
 GN Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92272714; PubMed-1590786;
 RA den Hertog J., Pals C.E., Jonk L.J., Krulder W.;
 RT "Differential expression of a novel murine non-receptor protein
 RT tyrosine phosphatase during differentiation of p19 embryonal
 RT carcinoma cells.";
 RL Biochem. Biophys. Res. Commun. 184:1241-1249(1992).
 RN [2]
 RP REVISIONS TO 297-416.
 RX MEDLINE-93112015; PubMed-1472029;
 RA Takekawa M., Itoh F., Hlnoda Y., Arimura Y., Toyota M., Sekiya M.,
 RA Adachi M., Imai K., Yachi A.;

```

RT      *Cloning and characterization of a human cDNA encoding a novel
RL      putative cytoplasmic protein-tyrosine-phosphatase.";
RM      Biochem. Biophys. Res. Commun. 189:123-1230(1992).
RN      (3)
RP      SEQUENCE FROM N.A.
RC      STRAIN-BALB/c;
RX      MEDLINE-95289971; PubMed-7772023;
RA      Charrest A., Wagner J., Shen S.H., Tremblay M.L.;
RT      "Murine protein tyrosine phosphatase-PEST, a stable cytosolic protein
RL      tyrosine phosphatase.";
CC      Biochem. J. 308:425-432(1995).
CC      -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC      tyrosine + phosphate.
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic
CC      -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC      TYROSINE PHOSPHATASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; X63440; CAA45037.1; ALT_SEQ.
DR      EMBL; X86781; CAA60477.1; -.
DR      HSSP; 006124; 2SHP.
DR      MGD; MGI:104673; Pcpn12.
DR      InterPro; IPR000387; TYR_phosphatase.
DR      InterPro; IPR000242; TYR_PP.
DR      Pfam; PF00102; Y_phosphatase; 1.
DR      PRINTS; PR00700; PRTPHPTASE.
DR      SMART; SM00194; PTPc; 1.
DR      PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR      PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
DR      PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW      Hydrolyase.
FT      DOMAIN 27 291 PROTEIN-TYROSINE PHOSPHATASE.
FT      ACT_SITE 231 231 K -> N (IN REF. 1).
FT      CONFLICT 296 296 K -> N (IN REF. 1).
FT      CONFLICT 328 332 KODSP -> DETS (IN REF. 1).
FT      CONFLICT 380 380 W -> V (IN REF. 1).
SQ      SEQUENCE 775 AA; 86992 MW; 7106D73F5014E411 CRC64;

```

```

Query Match 19.2%; Score 71; DB 1; Length 775;
Best Local Similarity 57.1%; Pred. No. 0.55;
Matches 12; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
QY      18 SGNQPSNREKNRYDILPFOH 38
       :|:|:|||||
Db      52 TGEKENYKKNRYDILPFDH 72

```

Search completed: June 6, 2003, 11:29:29
 Job time : 11 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: June 6, 2003, 11:27:36 ; Search time 29 Seconds
(without alignments)
476,040 Million cell updates/sec

Title: US-09-935-703-7

Perfect score: 369
Sequence: 1 MOEFMALELKNLPGEFNSGN.....TTFWHSNGAVSLRLRYCA 67

Scoring table: BLOSUM62
Gapop 10.0 , gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rhiz:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score distribution, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	180	48.8	398	4	O9Y406 mus sapien
2	121	32.8	426	11	O55082 mus musculu
3	88.5	24.0	297	11	O63476 ratu
4	88.5	24.0	536	4	O8TE48 homo sapien
5	88.5	24.0	642	4	O96K06 mus musculu
6	88.5	24.0	642	11	O60986 mus musculu
7	88.5	24.0	659	11	O63477 ratu
8	88.5	24.0	659	11	O61042 mus musculu
9	88.5	24.0	536	11	O35128 mus musculu
10	88.5	24.0	595	11	O92163 mus musculu
11	88.5	24.0	597	11	O9R0V6 mus musculu
12	86	23.3	597	4	O969V8 homo sapien
13	86	23.3	613	11	P81718 ratu
14	86	23.3	624	4	O9UK67 homo sapien
15	85	23.0	827	5	O21055 caenorhabd
16	83	22.5	595	13	O92124 xenopus lae

17	79.5	21.5	793	11	O91V35 mus musculu
18	79.5	21.5	802	4	O96TD9 homo sapien
19	79	21.4	1064	5	O8SXB2 drosophila
20	79	21.4	1767	5	O9W4F5 drosophila
21	79	21.4	1767	5	O24495 drosophila
22	77.5	21.0	956	17	O8TPY5 methanosarc
23	76.5	20.7	289	5	O9N4M9 caenorhabd
24	75.5	20.5	833	13	O9DES7 brachydanio
25	75	20.3	1647	5	O9YVW1 drosophila
26	73.5	19.9	807	13	O91969 gallus gall
27	73.5	19.9	1122	11	O8R388 mus musculu
28	72.5	19.6	1117	11	O8R388 mus musculu
29	72	19.5	274	11	O9QW08 ratu
30	72	19.5	460	4	O96HD7 homo sapien
31	72	19.5	550	5	O44180 caenorhabd
32	72	19.5	593	13	O90687 gallus gall
33	72	19.5	597	11	O64509 mus musculu
34	71	19.2	382	11	O63745 ratu
35	71	19.2	433	13	O9PT91 brachydanio
36	70	19.0	353	4	O96AUS homo sapien
37	70	18.8	406	11	O922E7 mus musculu
38	69.5	18.8	1437	5	O44329 hiru
39	69	18.7	336	11	O9QWQ7 mus musculu
40	69	18.7	1114	4	O9H0Y6 homo sapien
41	69	18.7	1143	4	O16614 homo sapien
42	69	18.7	1291	11	O61812 mus musculu
43	69	18.7	1343	11	O64730 mus musculu
44	68.5	18.6	822	13	O91556 xenopus lae
45	68.5	18.6	1200	13	O91054 heterodontu

ALIGNMENTS

RESULT 1

ID O9Y406 PRELIMINARY; PRT; 398 AA.

AC O9Y406; 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DR 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE Hypothetical 45.7 kDa protein (Fragment).
GN DKFZ566K0524.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Ansoorge W., Wilkner U., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL050040; CAB43248.1; -
DR HSSP; O06124; 2SHP.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; TYR_PP.
DR Pfam; PF00102; 1-phosphatase. 1.
DR PRINTS; PR00700; PRTYPPHASE.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydroxylase; Hypothetical protein.
FT NON_TER
SQ SEQUENCE 398 AA; 45690 MW; 857AAD03747870A2 CRC64;

Query Match 48.8%; Score 180; DB 4; Length 398;
Best Local Similarity 94.4%; Pred. No. 1.6e-14;
Matches 34; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 MOEFMALELKNLPGEFNSGNOPSNREKRRYDILPF 36
DB 138 MOEFMALELKNLPGEFNSGNOPSNREKRRYDILPF 173


```

RESULT 2
055082 ID 055082 PRELIMINARY; PRT; 426 AA.
AC 055082;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Protein-tyrosine-phosphatase (EC 3.1.3.48).
GN PTPN20.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS.
RX MEDLINE=98070510; PubMed=9407093;
RA Ohnogi M., Kuramochi S., Matsuda S., Yamamoto T.;
RT "Molecular cloning and characterization of a novel cytoplasmic
RT protein-tyrosine phosphatase that is specifically expressed in
RL spermatoocytes."
DR J. Biol. Chem. 272:33092-33099(1997).
DR EMBL: D64141; BAA23761.1; -.
DR HSSP; Q06124; ZSPH.
DR MGI; MGI:1196295; Pcpn20.
DR InterPro: IPR000387; Pcpn20.
DR InterPro: IPR000242; Tyr_PP.
DR Pfam: PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPc; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
SQ Hydrolase.
SQ SEQUENCE 426 AA; 49118 MW; 2B35FB13379502F4 CRC64;

Query Match 32.8%; Score 121; DB 11; Length 426;
Best Local Similarity 61.1%; Pred. No. 5,6e-07;
Matches 22; Conservative 6; Mismatches 8; Indels 0; Gaps 0.

OY 1 MOEFALTELNKNGEFGNSGNSNNEKNRYRDLTP 36
DB 166 IREFLEQMTLPDDFNSGNTLQNDKDKRYRDLTP 201
::||: || ||:||||| ||:|||||||:

RESULT 3
063476 ID 063476 PRELIMINARY; PRT; 297 AA.
AC 063476;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Protein tyrosine phosphatase epsilon M precursor (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN.
RX MEDLINE=96158952; PubMed=8579581;
RA Nakamura K., Mizuno Y., Kikuchi K.;
RT "Molecular cloning of a novel cytoplasmic protein tyrosine phosphatase
RT p19psep10n."
DR Biochem. Biophys. Res. Commun. 218:726-732(1996).
DR EMBL: D78613; BAA11433.1; -.
DR HSSP; P18052; IYFO.
DR InterPro: IPR000242; Tyr_PP.
DR Pfam: PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPc; 1.

```

```

DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
RW SIGNAL.
FT SIGNAL.
FT NON_TER
SQ SEQUENCE 297 AA; 18 POTENTIAL.
FT 297
SQ SEQUENCE 297 AA; 34221 MW; BCPAA633EF6526DF CRC64;

Query Match
Best Local Similarity 24.0%; Score 88.5; DB 11; Length 297;
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3

QY 2 QEFNALELKNLPGEFNSGNQPSNRKKNRYDILPROH-----H 39
DB 136 EEFNSLPSGHIOGFPELLANKRENEKNKRYPIILPNDRHCVILSQLDGI PCSDIYINASYID 195
QY 40 GY-----SGNERTT--FW 51
DB 196 GYKKNKFIAAGPKQETVNDWF 218

RESULT 4
Q8TE48 PRELIMINARY; PRT; 536 AA.
AC Q8TE48;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Tyrosine phosphatase epsilon PDI.
GN PTPRE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Wabacken T., Hauge H., Funderud S., Aasheim H.C.;
RT "Characterisation, expression and functional aspects of a novel
RT protein tyrosine phosphatase epsilon isoform."
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
SQ EMBL, AJ430580; CAD23182.1;
SQ SEQUENCE 536 AA; 60799 MW; 06C29C8CACCE6A CRC64;

Query Match
Best Local Similarity 24.0%; Score 88.5; DB 4; Length 536;
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

QY 2 QEFNALELKNLPGEFNSGNQPSNRKKNRYDILPROH-----H 39
DB 79 EEFNSLPSGHIOGFPELLANKRENEKNKRYPIILPNDRHCVILSQLDGI PCSDIYINASYID 138
QY 40 GY-----SGNERTT--FW 51
DB 139 GYKKNKFIAAGPKQETVNDWF 161

RESULT 5
Q96K06 PRELIMINARY; PRT; 642 AA.
AC Q96K06;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Tyrosine phosphatase epsilon.
GN PTPRE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Wabacken T.K., Hauge H., Finne E.F., Wiedlocha A., Aasheim H.C.;
RT "Expression of human protein tyrosine phosphatase epsilon in
RT leukocytes, a potential ERK-pathway regulating phosphatase.",
RT Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases

```

DR EMBL: AJ31569; CAC86583.1; -.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR InterPro: IPR000242; TYR_PP.
 DR Pfam: PF00102; Y_phosphatase.2.
 DR PROSITE: PS00383; TYR_PHOSPHATASE.1; UNKNOWN_2.
 DR PROSITE: PS50056; TYR_PHOSPHATASE.2; 2.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 2.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 2.
 SO SEQUENCE 642 AA; 74581 MW; D32A288A1866D948 CRC64;

Query Match 24.0%; Score 88.5; DB 4; Length 642;
 Best Local Similarity 28.9%; Pred. No. 0.012;
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

OY 2 QEFMALEKLNLPGEFNSGNQPSNREKNRYRDILPFQH-----H 39
 DB 79 EEFNLSPGHIOGTFLANKKEENREKNRYPNILPNDHCRVILSQDGIQPCSDYINASYID 138

OY 40 GY-----SGPNERT--FW 51
 DB 139 GYKEKNKFLAAGPKQETVNDWF 161

RESULT 6

ID 060986 PRELIMINARY; PRT; 642 AA.
 AC 060986;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Epsilon tyrosine phosphatase cytoplasmic isoform.
 GN PTPRE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_Taxid-10090;
 RX MEDLINE-96109240; PubMed-8618876;
 RA Elson A., Leder P.;
 RT "Identification of a cytoplasmic, phorbol ester-inducible isoform of
 protein tyrosine phosphatase epsilon.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:12235-12239(1995).
 DR EMBL: U36758; AAC52331.1; -.
 DR HSSP: P18052; IYFO.
 DR MGD: MGI:97813; PTPre.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR InterPro: IPR000242; TYR_PP.
 DR Pfam: PF00102; Y_phosphatase.2.
 DR PRINTS: PR00700; PRTYPHPTASE.
 DR SMART: SM00194; PTPc; 2.
 DR PROSITE: PS00383; TYR_PHOSPHATASE.1; 2.
 DR PROSITE: PS50056; TYR_PHOSPHATASE.2; 2.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 2.
 KW Hydrolyase.
 SO SEQUENCE 642 AA; 74674 MW; EB89E78259C60A47 CRC64;

Query Match 24.0%; Score 88.5; DB 11; Length 642;
 Best Local Similarity 28.9%; Pred. No. 0.012;
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

OY 2 QEFMALEKLNLPGEFNSGNQPSNREKNRYRDILPFQH-----H 39
 DB 79 EEFNLSPGHIOGTFLANKKEENREKNRYPNILPNDHCRVILSQDGIQPCSDYINASYID 138

OY 40 GY-----SGPNERT--FW 51
 DB 139 GYKEKNKFLAAGPKQETVNDWF 161

RESULT 7

ID 063477 PRELIMINARY; PRT; 659 AA.
 AC 063477;

DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Protein tyrosine phosphatase epsilon C (fragment).
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NC NCB1_Taxid-10116;
 RX MEDLINE-96158952; PubMed-8579581;
 RA Nakamura K., Mizuno Y., Kikuchi K.;
 RT "Molecular cloning of a novel cytoplasmic protein tyrosine phosphatase
 PTPepsilon.";
 RL Blochem. Biophys. Res. Commun. 218:736-732(1996).
 DR EMBL: D78610; BAA20333.1; -.
 DR HSSP: P18052; IYFO.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR InterPro: IPR000242; TYR_PP.
 DR Pfam: PF00102; Y_phosphatase.2.
 DR PRINTS: PR00700; PRTYPHPTASE.
 DR SMART: SM00194; PTPc; 2.
 DR PROSITE: PS00383; TYR_PHOSPHATASE.1; 2.
 DR PROSITE: PS50056; TYR_PHOSPHATASE.2; 2.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 2.
 KW Hydrolyase.
 FT NON_TER
 SO SEQUENCE 659 AA; 76451 MW; FCC459BB5DD11A4 CRC64;

Query Match 24.0%; Score 88.5; DB 11; Length 659;
 Best Local Similarity 28.9%; Pred. No. 0.012;
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

OY 2 QEFMALEKLNLPGEFNSGNQPSNREKNRYRDILPFQH-----H 39
 DB 96 EEFNLSPGHIOGTFLANKKEENREKNRYPNILPNDHCRVILSQDGIQPCSDYINASYID 155

OY 40 GY-----SGPNERT--FW 51
 DB 156 GYKEKNKFLAAGPKQETVNDWF 178

RESULT 8

ID 061042 PRELIMINARY; PRT; 699 AA.
 AC 061042;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Protein tyrosine phosphatase.
 GN PTPRE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_Taxid-10090;
 RX MEDLINE-96181534; PubMed-8610169;
 RA Schmidt A., Rutledge S.J., Endo N., Ogas E., Tanaka H., Wesolowski G.,
 Leu C.T., Huang Z., Ramachandaran C., Rodan S.B., Rodan G.A.;
 RT "Protein-tyrosine phosphatase activity regulates osteoclast formation
 and function: inhibition by alendronate.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:3068-3073(1996).
 DR EMBL: U40280; AAB02190.1; -.
 DR HSSP: P18052; IYFO.
 DR MGD: MGI:97813; PTPre.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR InterPro: IPR000242; TYR_PP.

Query Match 24.0%; Score 88.5; DB 11; Length 699;
 Best Local Similarity 28.9%; Pred. No. 0.012;
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

OY 2 QEFMALEKLNLPGEFNSGNQPSNREKNRYRDILPFQH-----H 39
 DB 96 EEFNLSPGHIOGTFLANKKEENREKNRYPNILPNDHCRVILSQDGIQPCSDYINASYID 155

DR Pfam: PF00102; Y_phosphatase; 2.
 DR PRINTS: PR00700; PRTPHPTASE.
 DR SMART: SM00194; PTPc; 2.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 2.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 2.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 2.
 DR KMW_Hydrolase.
 SQ SEQUENCE 699 AA; 80687 MW; 581F9CB881BC05B CRC64;
 Query Match 24.0%; Score 88.5; DB 11; Length 699;
 Best Local Similarity 28.9%; Pred. No. 0.013;
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;
 OY 2 QEFMAL-ELKNGEFGNSGNPSNRKRRDILPFQH-----H 39
 DB 136 EEFNSLQKQEVKNLHORL-EGGRPEKSKNRYKNILPFQH 195
 OY 40 GY-----SGPNERT--FW 51
 DB 196 GYKRNKFTIAAGPKQETVNDPW 218

RESULT 9

ID 035128 PRELIMINARY; PRT; 556 AA.
 AC 035128;
 DT 01-JAN-1998 (TREMBLrel. 05 Created)
 DT 01-JAN-1998 (TREMBLrel. 05; Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21; Last annotation update)
 DE PPN6.
 GN PPN6.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98112780; PubMed=9445485;
 RA Ansari-Lari M.A., Oeltjen J.C., Schwartz S., Zhang Z., Muzny D.M.,
 Lu J., Gorrell J.H., Chinault A.C., Belmont J.W., Miller W.,
 Gibbs R.A.;
 RT "Comparative sequence analysis of a gene-rich cluster at human
 chromosome 12p13 and its syntenic region in mouse chromosome 6.";
 RL Genome Res. 8:29-40(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX Muzny D., Ansari-Lari M.A., Timms K.M., Yu W., Dugan S., Lu J.,
 Shen Y., Rowland K., Liu W., Perez L., Ding Y., Haywood M., Jain A.,
 Leal B., Logan O., Nguyen V., Savage L., Shen H., Worley K., Chen E.,
 Forcum J., Atkinson A.D., Chiu M.W., Gorrell J.H., Brundage E., Di W.,
 Chinault C., Nelson D., Gibbs R.A.;
 RT "Direct Submission."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX Muzny D., Atkinson A.D., Adams C., Brundage E., Bunac C., Carvelli K.,
 Chacko J., Chen J., Di W., Ding Y., Dugan S., Durbin J., Forcum J.,
 Ganesh R., Garcia C., Goodman M., Gorrell J.H., Haywood M.,
 Hernandez J., Jackson L., Jin S., Kampel R., Karpaty S., Kovar C.,
 Leal B., Li Y., Licharge O., Liu W., Logan O., Lu J., Ly T.,
 Martinez C., Oswal G., Perez L., Rashid N.D., Rowland K., Savage L.,
 Scherer S.E., Shen H., Simon M., Stovall K., Timms K.M., Todd J.,
 Vo O., Williamson A., Worley K.C., Yu W., Chinault C., Nelson D.,
 Gibbs R.A.;
 RT "Direct Submission."
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR HSSP: P29350; 1GWZ.
 DR InterPro: IPR000980; SH2.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR Pfam: PF00017; SH2; 2.
 DR Pfam: PF00102; Y_phosphatase; 1.

DR PRINTS: PR00700; PRTPHPTASE.
 DR Prodom: PD000093; SH2; 2.
 DR SMART: SM00194; PTPc; 1.
 DR SMART: SM00252; SH2; 1.
 DR PROSITE: PS50001; SH2; 2.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
 DR KMW_Hydrolase.
 SQ SEQUENCE 556 AA; 63179 MW; CACF025DF9D7BC2D CRC64;
 Query Match 23.8%; Score 88; DB 11; Length 556;
 Best Local Similarity 47.5%; Pred. No. 0.012;
 Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;
 OY 2 QEFMAL-ELKNGEFGNSGNPSNRKRRDILPFQH 38
 DB 207 EEFNSLQKQEVKNLHORL-EGGRPEKSKNRYKNILPFQH 245

RESULT 10

ID 0921G3 PRELIMINARY; PRT; 595 AA.
 AC 0921G3;
 DT 01-DEC-2001 (TREMBLrel. 19 Created)
 DT 01-DEC-2001 (TREMBLrel. 19; Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21; Last annotation update)
 DE Hemopoietic cell phosphatase.
 GN HCPH.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX Strausberg R.;
 RA Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC012660; AAH12660.1;
 DR MGI: 96055; HCPH.
 DR InterPro: IPR000980; SH2.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR Pfam: PF00102; Y_phosphatase; 1.
 DR Prodom: PD000093; SH2; 2.
 DR PROSITE: PS50001; SH2; 2.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; UNKNOWN_1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
 SQ SEQUENCE 595 AA; 67545 MW; CFI6750D032638D2 CRC64;
 Query Match 23.8%; Score 88; DB 11; Length 595;
 Best Local Similarity 47.5%; Pred. No. 0.013;
 Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;
 OY 2 QEFMAL-ELKNGEFGNSGNPSNRKRRDILPFQH 38
 DB 246 EEFNSLQKQEVKNLHORL-EGGRPEKSKNRYKNILPFQH 284

RESULT 11

ID 09ROV6 PRELIMINARY; PRT; 597 AA.
 AC 09ROV6;
 DT 01-MAY-2000 (TREMBLrel. 13 Created)
 DT 01-MAY-2000 (TREMBLrel. 13; Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21; Last annotation update)
 DE SH2 phosphatase 1.
 GN SHP-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C3H;
RA Tsui H.W., Tsui F.;
RT "Murine SHP-1";
RL Nat. Genet. 0:0-0(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C3H;
RA MEDLINE-9348302; PubMed-10419485;
RA Martin A., Tsui H.W., Shulman M.J., Iseman D., Tsui F.W.;
RT "Murine SHP-1 splice variants with altered Src homology 2 (SH2)
RT domains. Implications for the SH2-mediated intramolecular regulation
RT of SHP-1."
RL J. Biol. Chem. 274:21725-21734(1999).
DR EMBL; U65955; AAD00151.1; -.
DR EMBL; U65951; AAD00151.1; JOINED.
DR EMBL; U65952; AAD00151.1; JOINED.
DR EMBL; U65953; AAD00151.1; JOINED.
DR EMBL; U65954; AAD00151.1; JOINED.
DR HSSP; P29350; IGMZ.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; TYR_PP.
DR Pfam; PF00017; SH2; 2.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRYPPHPTASE.
DR PRINTS; PR00401; SH2DOMAIN.
DR ProDom; PD000093; SH2; 2.
DR SMART; SM00194; PTPc; 1.
DR SMART; SM00252; SH2; 2.
DR PROSITE; PS50001; SH2; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydroxylase.
SQ SEQUENCE 597 AA; 67717 MW; E8491CE77E06E989 CRC64;

Query Match 23.8%; Score 88; DB 11; Length 597;
Best Local Similarity 47.5%; Pred. No. 0.013;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

QY 2 QEFMAL---ELKNLPGEFNSGNPSNEKNRYDILPFQ 38
DB 248 EEFESLQKQEVKNLHQL-EGQREPKSKNNKYLDPDH 286

RESULT 12
ID Q969V8 PRELIMINARY; PRT; 597 AA.
AC Q969V8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Protein tyrosine phosphatase, non-receptor type 6.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC007667; AAH07667.1; -.
DR EMBL; BC007523; AAH07523.1; -.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR000387; TYR_phosphatase.

```

```

DR InterPro: IPR000242; TYR_PP.
DR Pfam; PF00017; SH2; 2.
DR Pfam; PF00102; Y_phosphatase; 1.
DR ProDom; PD000093; SH2; 2.
DR PROSITE; PS50001; SH2; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; UNKNOWN_1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Receptor.
SQ SEQUENCE 597 AA; 67719 MW; 6A291A2860159389 CRC64;

Query Match 23.3%; Score 86; DB 4; Length 597;
Best Local Similarity 47.5%; Pred. No. 0.023;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

QY 2 QEFMAL---ELKNLPGEFNSGNPSNEKNRYDILPFQ 38
DB 248 EEFESLQKQEVKNLHQL-EGQREPKSKNNKYLDPDH 286

RESULT 13
ID P81718 PRELIMINARY; PRT; 613 AA.
AC P81718;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Protein-tyrosine phosphatase N6 (EC 3.1.3.48).
GN SHP-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Aoki N., Yamaguchi-Aoki Y., Ullrich A.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: PLAYS A KEY ROLE IN HEMATOPOIESIS. THIS PTPASE ACTIVITY
CC MAY DIRECTLY LINK GROWTH FACTOR RECEPTORS AND OTHER SIGNALING
CC PROTEINS THROUGH PROTEIN-TYROSINE PHOSPHORYLATION. THE SH2 REGIONS
CC MAY INTERACT WITH OTHER CELLULAR COMPONENTS TO MODULATE ITS OWN
CC PHOSPHATASE ACTIVITY AGAINST INTERACTING SUBSTRATES (BY
CC SIMILARITY).
CC - CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O -> PROTEIN
CC TYROSINE + ORTHOPHOSPHATE.
CC - SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC - SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.
CC - SIMILARITY: CONTAINS 2 SH2 DOMAINS.
DR EMBL; U77038; AAD00262.1; -.
DR HSSP; P29350; IGMZ.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; TYR_PP.
DR Pfam; PF00017; SH2; 2.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRYPPHPTASE.
DR PRINTS; PR00401; SH2DOMAIN.
DR ProDom; PD000093; SH2; 2.
DR SMART; SM00194; PTPc; 1.
DR SMART; SM00252; SH2; 2.
DR PROSITE; PS50001; SH2; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydroxylase; SH2 domain.
FT DOMAIN 6 102 SH2.
FT DOMAIN 112 215 SH2.
FT DOMAIN 271 516 PROTEIN-TYROSINE PHOSPHATASE.
FT ACT_SITE 455 455 BY SIMILARITY.
SQ SEQUENCE 613 AA; 69578 MW; 29364B28BF45C87 CRC64;

Query Match 23.3%; Score 86; DB 11; Length 613;

```

Best Local Similarity 47.5%; Pred. No. 0.024;
Matches 19; Conservative 6; Mismatches 11; Indels 4; Gaps 2;

OY 2 QEFMAL---ELKNLPGEFNSGNPSNREKNRYDILPFQH 38
DB 246 EEFESLQKQEVKNLHQR-EGORPENKSKNRYKNILPFQH 286

RESULT 14

ID 021055 PRELIMINARY; PRT: 624 AA.

AC 09UK67.
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 70 kDa SHP-1L protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99428497; PubMed-10497187;
RA Jin Y.J., Yu C.L., Burdoff S.J.,
RT "Human 70-kDa SHP-1L differs from 68-kDa SHP-1 in its C-terminal
structure and catalytic activity."
RL J. Biol. Chem. 274:28301-28307(1999).
DR EMBL: AF178946; AAD5317.1; -
DR HSSP: P29350; 1FPR.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; TYR_PP.
DR Pfam: PF00017; SH2; 2.
DR Pfam: PF00102; Y_phosphatase; 1.
DR PRINTS: PR00700; PRYPPHPTASE.
DR PRODOM: PD000093; SH2; 2.
DR SMART: SM00194; PTPc; 1.
DR SMART: SM00252; SH2; 2.
DR PROSITE: PS50001; SH2; 2.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolyase.
SQ SEQUENCE 624 AA: 70117 MW: CB73D7EC8396DDA0 CRC64;

Query Match 23.3%; Score 86; DB 4; Length 624;
Best Local Similarity 47.5%; Pred. No. 0.024;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

OY 2 QEFMAL---ELKNLPGEFNSGNPSNREKNRYDILPFQH 38
DB 246 EEFESLQKQEVKNLHQR-EGORPENKSKNRYKNILPFQH 284

RESULT 15

ID 021055 PRELIMINARY; PRT: 827 AA.

AC 021055; 019847;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE PTP-2 protein (EC 3.1.3.48)
OS P59G1.5 OR P28B12.4 OR PTP-2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Latreille P.;
RL Submitted (APR-1996) to the EMBL/Genbank/DBJ databases.

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
RN [3]

RP SEQUENCE OF 160-827 FROM N.A.
RA Gutch M.J., Tonks N.K., Hengartner M.O.;
RL Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: U53332; AKK3131.1; -
DR EMBL: AF015882; AAC21678.1; -
DR HSSP: Q06124; 2SHP.
DR InterPro: IPR002908; Frataxin_like.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; TYR_PP.
DR Pfam: PF01491; Frataxin_Cyay; 1.
DR Pfam: PF00017; SH2; 2.
DR Pfam: PF00102; Y_phosphatase; 1.
DR PRINTS: PR00700; PRYPPHPTASE.
DR PRODOM: PD000093; SH2; 2.
DR SMART: SM00646; Frataxin_like; 1.
DR SMART: SM00194; PTPc; 1.
DR SMART: SM00252; SH2; 2.
DR PROSITE: PS01344; FRATAXIN_1; UNKNOWN_1.
DR PROSITE: PS50810; FRATAXIN_2; 1.
DR PROSITE: PS50001; SH2; 2.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolyase.
SQ SEQUENCE 827 AA: 94923 MW: 1ED2DAD4C54024D6 CRC64;

Query Match 23.0%; Score 85; DB 5; Length 827;
Best Local Similarity 43.9%; Pred. No. 0.045;
Matches 18; Conservative 6; Mismatches 13; Indels 4; Gaps 1;

OY 2 QEFMALELKNLPGE---FNSGNPSNREKNRYDILPFQH 38
DB 425 EEFRLSQEALPAEQYLSKREGRPVNAEKRYKNIVPFQH 465

Search completed: June 6, 2003, 11:30:06
Job time: 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 11:28:11 ; Search time 15 Seconds

(without alignments)
429,400 Million cell updates/sec

Title: US-09-935-703-7

Perfect score: 1 MOEFMALELNKLPGEFNSGN.....TFMHGNSGAVSLIRYCA 67

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	180	48.8	398	2 T08716	protein-tyrosine-p
2	88.5	24.0	699	2 J06132	protein-tyrosine-p
3	88.5	24.0	700	1 S12053	protein-tyrosine-p
4	88	23.8	595	1 A44390	protein-tyrosine-p
5	86	23.0	595	1 S20825	protein-tyrosine-p
6	85	23.0	668	2 T34317	protein-tyrosine-p
7	83	22.5	595	1 A55651	protein-tyrosine-p
8	79.5	21.5	802	1 A36065	protein-tyrosine-p
9	79	21.4	1615	2 B49502	protein-tyrosine-p
10	79	21.4	1767	2 A49502	protein-tyrosine-p
11	77.5	21.0	796	1 J01285	protein-tyrosine-p
12	75	20.9	521	1 A44267	protein-tyrosine-p
13	75	20.3	1557	2 A41214	protein-tyrosine-p
14	75	20.3	1630	2 A41214	protein-tyrosine-p
15	73	19.8	841	2 A43254	protein-tyrosine-p
16	72	19.5	550	2 T32543	protein-tyrosine-p
17	72	19.5	585	2 A46209	protein-tyrosine-p
18	72	19.5	593	1 JN0805	protein-tyrosine-p
19	72	19.5	593	2 J05167	protein-tyrosine-p
20	72	19.5	597	1 A53593	protein-tyrosine-p
21	71	19.2	382	1 S48748	protein-tyrosine-p
22	71	19.2	773	1 JH0609	protein-tyrosine-p
23	71	19.2	775	1 S55345	protein-tyrosine-p
24	71	19.2	780	1 J01368	protein-tyrosine-p
25	71	19.2	802	1 B44390	protein-tyrosine-p
26	70	19.0	363	1 S14294	protein-tyrosine-p
27	70	19.0	382	1 A38191	protein-tyrosine-p
28	70	19.0	387	1 A60345	protein-tyrosine-p
29	70	19.0	415	1 A33899	protein-tyrosine-p

30	69.5	18.8	1437	2 T31093	probable protein-t
31	69	18.7	1175	2 S51005	protein-tyrosine-p
32	69	18.7	1176	2 S15845	protein-tyrosine-p
33	69	18.7	1291	1 A28334	protein-tyrosine-p
34	69	18.7	1304	1 A46546	leukocyte common a
35	68.5	18.6	989	2 S69711	hypothetical prote
36	68.5	18.6	1200	2 T43148	probable protein-t
37	68	18.4	432	1 A34845	protein-tyrosine-p
38	68	18.4	435	1 TPDH01	protein-tyrosine-p
39	68	18.4	1174	2 T38140	protein-tyrosine-p
40	68	18.4	1237	2 A54080	protein-tyrosine-p
41	67	18.2	356	2 TW0049	protein-tyrosine-p
42	67	18.2	829	1 A47373	protein-tyrosine-p
43	66.5	18.0	405	2 S68250	protein-tyrosine-p
44	66.5	18.0	405	2 S68250	protein-tyrosine-p
45	66.5	18.0	1187	1 A53661	protein-tyrosine-p

ALIGNMENTS

RESULT 1
T08716
protein-tyrosine-phosphatase homolog DKFZp566K0524.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 13-Aug-1999
C/Accession: T08716
R:Ansoorge, W.; Walchner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, May 1999
A:Reference number: Z16472
A:Accession: T08716
A:Molecule type: mRNA
A:Residues: 1-398 <ANS>
A:Cross-references: EMBL:AL050040
A:Experimental source: fetal kidney; clone DKFZp566K0524
C:Genetics:
A>Note: DKFZp566K0524.1
C:Superfamily: protein-tyrosine-phosphatase homology
F:101-379/Domain: protein-tyrosine-phosphatase homology <P>

Query Match 48.8%; Score 180; DB 2; Length 398;
Best Local Similarity 94.4%; Pred. No. 3.6e-14;
Matches 34; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 MOEFMALELNKLPGEFNSGNOPNREKRYDILPF 36
|||||
Db 138 MOEFMALELNKLPGEFNSGNOPNREKRYDILPF 173

RESULT 2
JC6132
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type epsilon precursor - mouse
N:Alternate names: phosphotyrosine phosphatase; protein-tyrosine-phosphate phosphohyd
C:Species: Mus musculus (house mouse)
C>Date: 16-Apr-1997 #sequence_revision 09-May-1997 #text_change 21-Jun-2002
R:Schmidt, A.; Rutledge, S.J.; Endo, N.; Opas, E.E.; Tanaka, H.; Wesolowski, G.; Leu,
Proc. Natl. Acad. Sci. U.S.A. 93, 3068-3073, 1996
A:Title: Protein-tyrosine phosphatase activity regulates osteoclast formation and fun
A:Reference number: JC6132; MUID:96181534; PMID:8610169
A:Contents: bone marrow cell
A:Accession: JC6132
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-699 <SCH>
A:Cross-references: GB:U40280; NID:91373052; PIDN:AAB02190.1; PID:91373053
C:Comment: This enzyme plays an important role in osteoclast formation and function 1
hosphonate action.
C:Genetics:
A:Gene: ptp
C:Superfamily: protein-tyrosine-phosphatase, receptor type alpha; leukocyte common an
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; transmembrane protein; ty
F:77-697/Domain: leukocyte common antigen cytosolic domain homology <LAC>

F:158-382/Domain: protein-tyrosine-phosphatase homology <PRP1>
 F:334/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:340/Binding site: substrate phosphatase (Arg) #status predicted
 F:629/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:635/Binding site: substrate phosphatase (Arg) #status predicted

Query Match 24.0% Score 88.5; DB 2; Length 699;
 Best Local Similarity 28.9%; Pred. No. 0.011;
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

QY 2 OEPMALINKLPGFNSGNQPSNRKRRYRDLIPFQH-----H 39
 DB 136 EEFNLSPSGHIOGTFFELANKKEENRKNRYNLPNDHSRVILSQVDGIPCSYINASYID 195
 QY 40 GY-----SGNERTT--FW 51
 DB 196 GYKERNKFTAAOGPKQETVNDWF 218

RESULT 3
 S12053
 protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type epsilon precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: S12053
 R:Kruenger, N.X.; Streuli, M.; Salto, H.
 EMBO J. 9, 3241-3252, 1990
 A:Title: Structural diversity and evolution of human receptor-like protein tyrosine phosphatases
 A:Reference number: S12049; MUID:91006018; PMID:2170109
 A:Accession: S12053
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-700 <NRD>
 A:Cross-references: GB:X54134; NID:935791; PIDN:CAA38069.1; PID:935792
 C:Genetics:
 A:Gene: GDB:PTPRP
 A:Cross-references: GDB:131385; OMIM:600926
 A:Map position: 10q26-10q26
 C:Function:
 A:Description: catalyzes the hydrolysis of peptidyl-tyrosine to release phosphate
 C:Superfamily: protein-tyrosine-phosphatase, receptor type alpha; leukocyte common antigen family; phosphoprotein; phosphoric monoester hydrolase; receptor; transmembrane protein-19/Domain: signal sequence #status predicted <STG>
 F:120-700/Product: protein-tyrosine-phosphatase, receptor type epsilon #status predicted
 F:47-63/Domain: transmembrane #status predicted <TMN>
 F:78-698/Domain: leukocyte common antigen cytosolic domain homology <LAC>
 F:159-383/Domain: protein-tyrosine-phosphatase homology <PRP1>
 F:335/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:341/Binding site: substrate phosphatase (Arg) #status predicted
 F:630/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:636/Binding site: substrate phosphatase (Arg) #status predicted

Query Match 24.0% Score 88.5; DB 1; Length 700;
 Best Local Similarity 26.9%; Pred. No. 0.011;
 Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

QY 2 OEPMALINKLPGFNSGNQPSNRKRRYRDLIPFQH-----H 39
 DB 137 EEFNLSPSGHIOGTFFELANKKEENRKNRYNLPNDHSRVILSQVDGIPCSYINASYID 196
 QY 40 GY-----SGNERTT--FW 51
 DB 197 GYKERNKFTAAOGPKQETVNDWF 219

RESULT 4
 A44390
 protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 6 - mouse
 N:Alternate names: hematopoietic cell phosphatase (HCP); protein-tyrosine-phosphatase 10
 C:Species: Mus musculus (house mouse)
 C:Date: 03-May-1994 #sequence_revision 19-May-1994 #text_change 11-Jun-1999
 C:Accession: A44390; F61180; I65741; I52816; I65740; A45143; B45143
 R:Matthews, R.J.; Bowne, D.B.; Flores, E.; Thomas, M.L.

MOL. Cell. Biol. 12, 2396-2405, 1992
 A:Title: Characterization of hematopoietic intracellular protein tyrosine phosphatase and threonine-rich sequences.
 A:Reference number: A44390; MUID:92236615; PMID:1373816
 A:Accession: A44390
 A:Molecule type: mRNA
 A:Residues: 1-595 <MAT>
 A:Cross-references: GB:M90389; NID:9200550; PIDN:AAA40007.1; PID:9200551
 R:Yl, T.L.; Cleveland, J.L.; Ihle, J.N.
 MOL. Cell. Biol. 12, 836-846, 1992
 A:Title: Protein tyrosine phosphatase containing SH2 domains: characterization, preference
 A:Reference number: A42031; MUID:92123209; PMID:1732748
 A:Accession: A42031
 A:Molecule type: mRNA
 A:Residues: 1-595 <YII>
 A:Cross-references: GB:M68902; NID:9193807
 A:Note: sequence extracted from NCBI backbone (NCBIN:76845, NCBI:76846)
 R:Yl, T.; Cleveland, J.L.; Ihle, J.N.
 Blood 78, 2222-2228, 1991
 A:Title: Identification of novel protein tyrosine phosphatases of hematopoietic cells
 A:Reference number: A61180; MUID:92032882; PMID:1932742
 A:Accession: F61180
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 342-451 <YIA>
 R:Shultz, L.D.; Schweitzer, P.A.; Rajan, T.V.; Yi, T.; Ihle, J.N.; Matthews, R.J.; Th
 Cell 73, 1445-1454, 1993
 A:Title: Mutations at the murine motheaten locus are within the hematopoietic cell pr
 A:Reference number: I52816; MUID:93313972; PMID:8324828
 A:Accession: I65741
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 334-353,359-382 <SHU1>
 A:Cross-references: GB:S63803; NID:9388449
 A:Note: deletion mutation
 A:Accession: I52816
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 350-358, 'EGGSPNLTPTPSSLVIVQYHTQ', 359-366 <SHU2>
 A:Cross-references: GB:S63763; NID:9388447
 A:Note: insertion mutation
 A:Accession: I65740
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 34-76, 'VPRPHNRAGGVTAAQGRALD' <SHU3>
 A:Cross-references: GB:S63764; NID:9388450
 A:Note: frameshift mutation
 R:Yung, Y.G.; Berg, K.L.; Pixley, F.J.; Angeletti, R.H.; Stanley, E.R.
 J. Biol. Chem. 267, 23447-23450, 1992
 A:Title: Protein tyrosine phosphatase-1C is rapidly phosphorylated in tyrosine in mac
 A:Reference number: A45143; MUID:93054686; PMID:1385421
 A:Accession: A45143
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 137-139, 'X', 141-143, 'X', 145-151 <YEU>
 A:Experimental source: BAC1.2P5 macrophage
 A:Note: sequence extracted from NCBI backbone (NCBIP:118519)
 A:Accession: B45143
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 54-56, 'X', 58, 'X', 60-61, 'X', 63-68 <YEU>
 A:Experimental source: BAC1.2P5 macrophage
 A:Note: sequence extracted from NCBI backbone (NCBIP:118518)
 C:Comment: This protein is found primarily in hematopoietic tissues.
 C:Genetics:
 A:Gene: me/HCPH; motheaten
 C:Function:
 A:Description: catalyzes hydrolysis of peptidyl-tyrosine to release phosphate
 C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-phosphatase
 C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase
 F:110-211/Domain: SH2 homology <SH2A>
 F:265-521/Domain: phosphatase catalytic domain #status predicted <PRP>

F:270-504/Domain: protein-tyrosine-phosphatase homology <PTP>
 F:453/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:459/Binding site: substrate phosphate (Arg) #status predicted

Query Match 23.8% Score 88; DB 1; Length 595;
 Best Local Similarity 47.5%; Pred. No. 0.011;
 Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

OY 2 QEFMAL---ELKNLPGEFNSGNOPSNRKNRYRDLIPROH 38
 Db 246 EEFESLQKQEVKNLHQR-EGOREPNKSKNRYKNIIPFDH 284

RESULT 5

Protein-tyrosine-phosphatase (EC 3.1.3.48) PTPN6, nonreceptor type 6 [validated] - human
 N/Alternate names: hematopoietic cell phosphatase HCP; protein-tyrosine-phosphatase 1C;
 C/Species: Homo sapiens (man)
 C/Date: 19-May-1994 #sequence-revision 08-Feb-1996 #text-change 21-Jun-2002
 R:Title: B42031; A38189; S20825; S17234; S20837
 R:Cell: Cleveland, J.L.; Inle, J.N.
 Moll. Cell. Biol. 12, 836-846, 1992
 A:Title: Protein tyrosine phosphatase containing SH2 domains: characterization, preferen
 A:Reference number: A42031; MUID:92123209; PMID:1732748
 A:Accession: B42031

A:Molecule type: mRNA
 A:Residues: 1-595 <Y11>

A:Cross-references: GB:M74093

A:Experimental source: T-lymphoid cell line

A:Note: sequence extracted from GenBank

R:Plutsky, J.; Neel, B.G.; Rosenberg, R.D.
 Proc. Natl. Acad. Sci. U.S.A. 89, 1123-1127, 1992

A:Title: Isolation of a src homology 2-containing tyrosine phosphatase.

A:Reference number: A38189; MUID:92141214; PMID:1736296

A:Accession: A38189

A:Molecule type: mRNA

A:Residues: 1-85; 'V', '87-595 <PLU>

A:Cross-references: GB:M77273; NID:q338079; PIDN:AAA36610.1; PID:q338080

A:Note: sequence extracted from NCBI database (NCBI:79619, NCBI:79620)

R:Shen, S.H.; Bastien, L.; Posner, B.I.; Chretien, P.
 Nature 333, 868, 1991

A:Title: Corrigendum: A protein-tyrosine phosphatase with sequence similarity to the SH2

A:Reference number: S20825

A:Accession: S20825

A:Molecule type: mRNA

A:Residues: 1; 'LSRG', '4-595 <SH2>

A:Cross-references: EMBL:X62055; NID:q35781; PIDN:CAA43982.1; PID:q35782

R:Shen, S.H.; Bastien, L.; Posner, B.I.; Chretien, P.
 Nature 352, 736-739, 1991

A:Title: A protein-tyrosine phosphatase with sequence similarity to the SH2 domain of th

A:Reference number: S17234; MUID:91343005; PMID:1652101

A:Accession: S17234

A:Molecule type: mRNA

A:Residues: 1; 'LSRG', '4-589', 'VPSRGSERCCPQVAMPQ' <SH2>

A:Experimental source: breast carcinoma cells

A:Note: sequence revised in reference S20805

C:Genetics:

A:Gene: GDB:PTPN6

A:Cross-references: GDB:131389; OMIM:176883

A:Map position: 12p13-12p13

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-phosphatase

F:4-98/Domain: SH2 homology <SH2>

F:110-211/Domain: SH2 homology <SH2>

F:265-521/Domain: phosphatase catalytic domain #status predicted <PHP>

F:270-504/Domain: protein-tyrosine-phosphatase #status predicted <PTP>

F:453/Active site: Cys (phosphocysteine intermediate) #status predicted

F:459/Binding site: substrate phosphate (Arg) #status predicted

Query Match 23.3% Score 86; DB 1; Length 595;
 Best Local Similarity 47.5%; Pred. No. 0.019;
 Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

OY 2 QEFMAL---ELKNLPGEFNSGNOPSNRKNRYRDLIPROH 38
 Db 246 EEFESLQKQEVKNLHQR-EGOREPNKSKNRYKNIIPFDH 284

RESULT 6

Protein-tyrosine-phosphatase (EC 3.1.3.48) PTP-2, nonreceptor type 6/11 [validated] -
 C/Species: Caenorhabditis elegans
 C/Date: 29-Oct-1999 #sequence-revision 29-Oct-1999 #text-change 30-Jun-2002
 C/Accession: T34317; T16193; T42211
 R:Latreille, P.

submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid F59G1.

A:Reference number: Z21504

A:Accession: T34317

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-668 <LAT>

A:Cross-references: EMBL:U53332; PIDN:MAC71160.1; GSPDB:GN00020; CESP:F59G1.5

A:Experimental source: strain Bristol N2; clone F59G1

R:Latreille, P.
 submitted to the EMBL Data Library, June 1995

A:Description: The sequence of C. elegans cosmid F28B12.

A:Reference number: Z18474

A:Accession: T16193

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-668 <LAT>

A:Cross-references: EMBL:U29537; NID:9868268; PID:9868272; PIDN:AAA68799.1; CESP:F28B

A:Experimental source: strain Bristol N2

R:Guth, M.J.; Flint, A.J.; Keller, J.; Tonks, N.K.; Hengartner, M.O.
 Genes Dev. 12, 571-585, 1998

A:Title: The Caenorhabditis elegans SH2 domain-containing protein tyrosine phosphatase

A:Reference number: Z22076; MUID:98139525; PMID:9472025

A:Accession: T42211

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 1-668 <GUP>

A:Cross-references: EMBL:AF015882; NID:92323369; PIDN:MAC21678.1; PID:92323370

C:Genetics:

A:Gene: ptp-2; F59G1.5; F28B12.4

A:Map position: 2

A:Introns: 7/2; 230/3; 269/3; 586/3

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-pho

C:Keywords: phosphoric monoester hydrolase

Query Match 23.0% Score 85; DB 2; Length 668;
 Best Local Similarity 43.9%; Pred. No. 0.029;
 Matches 18; Conservative 6; Mismatches 13; Indels 4; Gaps 1;

OY 2 QEFMALDKNLPGE---FNSGNOPSNRKNRYRDLIPROH 38
 Db 266 EEFRLSQEALPAEOYLTKREGRRVNAEKRYKNIIPFDH 306

Protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 11 - African clawed frog

N/Alternate names: SH-PTP2

C/Species: Xenopus laevis (African clawed frog)

C/Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 10-Sep-1999

C/Accession: A55651

R:Tang, T.L.; Freeman, Jr., R.M.; O'Reilly, A.M.; Neel, B.G.; Sokol, S.Y.
 Cell 80, 473-483, 1995

A:Title: The SH2-containing protein-tyrosine phosphatase SH-PTP2 is required upstream

A:Reference number: A55651; MUID:95163101; PMID:7859288

A:Accession: A55651

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-595 <TAN>

A:Cross-references: GB:U15287; NID:9601781; PIDN:AAA65731.1; PID:9601782

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-pho

C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase
 F:6-100/Domain: SH2 homology <SH2A>
 F:112-214/Domain: SH2 homology <SH2B>
 F:273-510/Domain: protein-tyrosine-phosphatase homology <PTP>
 F:459/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:465/Binding site: substrate phosphate (Arg) #status predicted

Query Match 22.5%; Score 83; DB 1; Length 595;
 Best Local Similarity 40.5%; Pred. No. 0.044;
 Matches 17; Conservative 8; Mismatches 11; Indels 6; Gaps 2;

QY 19 GNOPSREKKNRYDILPPOH-----HGSGPNERFTFMHGSN 55
 DB 268 GORENNKKNRYKILPDRTRVELHD-GDPNEQVSDYINAN 308

RESULT 8
 A36065

protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type alpha precursor - human
 C:Species: Homo sapiens (man)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

A:Accession: A36065; S17371; S12905; S13085
 R:Kaplan, R.; Morse, B.; Hübner, K.; Croce, C.; Hawk, R.; Ravera, M.; Ricca, G.; Jaye, M.; Natl. Acad. Sci. U.S.A. 87, 7000-7004, 1990
 A:Title: Cloning of three human tyrosine phosphatases reveals a multigene family of receptors
 A:Reference number: A36065; MUID:90384936; PMID:2169617
 A:Accession: A36065

A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-802 <KAP>

A:Cross-references: GB:A34668; NID:9190738; PIDN:AAA36528.1; PID:9190739
 R:Kruenger, M.A.; Stenull, M.; Salto, H.
 EMBL J. 9, 3241-3252, 1990

A:Title: Structural diversity and evolution of human receptor-like protein tyrosine phosphatases
 A:Reference number: S12049; MUID:9106018; PMID:2170109
 A:Accession: S12049

A:Status: preliminary
 A:Molecule type: mRNA

A:Residues: 1-138, 148-802 <KR0>
 A:Cross-references: GB:A54130; NID:935785; PIDN:CAA38065.1; PID:935786
 R:Jirik, F.R.; Janzen, N.M.; Melhado, I.G.; Harder, K.W.; Leslie, K.B.

A:Title: Cloning and chromosomal assignment of a widely expressed human receptor-like protein
 A:Reference number: S12905; MUID:91032191; PMID:2172030
 A:Accession: S12905

A:Status: preliminary
 A:Molecule type: mRNA

A:Residues: 1-121, 'P', 123-138, 148-199, 'G', 201-203, 'C', 205-802 <JIR>
 A:Cross-references: EMBL:X53364
 R:Jirik, F.R.; Janzen, N.M.; Melhado, I.G.; Harder, K.W.

A:Title: Cloning and chromosomal assignment of a widely expressed human receptor-like protein
 A:Reference number: S12905; MUID:91032191; PMID:2172030
 A:Accession: S12905

A:Status: preliminary
 A:Molecule type: mRNA

A:Residues: 1-61, 'V', 62-82, 84-121, 'P', 123-138, 148-199, 'G', 201-203, 'C', 205-802 <JIR>
 R:Ohagi, S.; Nishii, M.; Steiner, D.F.

A:Title: Sequence of a cDNA encoding human LRP (leukocyte common antigen-related peptide)
 A:Reference number: S13085; MUID:91088320; PMID:2175890
 A:Accession: S13085

A:Status: preliminary
 A:Molecule type: mRNA

A:Residues: 1-113, 'N', 115-138, 148-288, 'E', 290-366, 'A', 368-492, 'S', 494-785, 'E', 787-802 <C>
 A:Cross-references: EMBL:X54890; NID:932312; PIDN:CAA38662.1; PID:932313
 C:Genetics:

A:Gene: GDB:PTPA; PTPA
 A:Cross-references: GDB:126732; OMIM:176884
 A:Map position: 20p13-20p13

C:Function: catalyzes the hydrolysis of peptidyl-phosphotyrosine to release phosphate
 A:Note: allows recovery from insulin stimulation by dephosphorylating insulin receptor

C:Superfamily: protein-tyrosine-phosphatase, receptor type alpha; leukocyte common antigen
 C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; rec
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-802/Product: protein-tyrosine-phosphatase, receptor type alpha #status predicted
 F:20-154/Domain: extracellular #status predicted <EXT>
 F:155-171/Domain: transmembrane #status predicted <TMN>
 F:172-802/Domain: intracellular #status predicted <INT>
 F:184-800/Domain: leukocyte common antigen cytosolic domain homology <LAC>
 F:285-490/Domain: protein-tyrosine-phosphatase homology <PTP1>
 F:558-780/Domain: protein-tyrosine-phosphatase homology <PTP2>
 F:442/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:448/Binding site: substrate phosphate (Arg) #status predicted
 F:732/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:738/Binding site: substrate phosphate (Arg) #status predicted

Query Match 21.5%; Score 79.5; DB 1; Length 802;
 Best Local Similarity 26.5%; Pred. No. 0.17;
 Matches 22; Conservative 9; Mismatches 19; Indels 33; Gaps 3;

QY 2 OEFMALKLNLPDEFNSGNPSREKKNRYDILPPOH-----H 39
 DB 243 EEFNALPACPIQATCEASKEENKKNRYNLLPYDSRVHLTPVEGVPDSYINAFIN 302

QY 40 GY-----SGPNERFT--FW 51
 DB 303 GYOEKKNKFLAAGPKETVNDFW 325

RESULT 9
 B49502

protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type 4E, splice form B precursor
 C:Species: Drosophila melanogaster
 C>Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 24-Apr-1998

A:Accession: B49502
 R:Oon, S.H.; Hong, A.; Yang, X.; Chia, W.

A:Title: Alternative splicing in a novel tyrosine phosphatase gene (DPTP4E) of Drosophila
 A:Reference number: A49502; MUID:94043220; PMID:8226938
 A:Accession: B49502

A:Molecule type: mRNA
 A:Residues: 1-1615 <CON>

A:Cross-references: GB:I120894
 C:Genetics:

A:Gene: FlyBase:PTP4E
 A:Cross-references: FlyBase:Fbgn0004368
 A:Introns: 1605/3

C:Superfamily: protein-tyrosine-phosphatase, receptor type 4E; fibronectin type III r
 C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; rec
 F:1254-1270/Domain: transmembrane #status predicted <TMN>
 F:1271-1615/Domain: intracellular #status predicted <INT>
 F:1353-1573/Domain: protein-tyrosine-phosphatase homology <PTP1>
 F:1525/Active site: Cys (phosphocysteine intermediate) #status predicted
 F:1531/Binding site: substrate phosphate (Arg) #status predicted

Query Match 21.4%; Score 79; DB 2; Length 1615;
 Best Local Similarity 25.3%; Pred. No. 0.44;
 Matches 25; Conservative 10; Mismatches 28; Indels 36; Gaps 3;

QY 2 OEFMALKLNLPDEFNSGNPSREKKNRYDILPPOH-----H 39
 DB 1331 EEFELKLVHGRDQACFAALPCNRPKNRTNLLPYDSRFKLPVDDGSDYINAYMP 1390

QY 40 GYSGPNE-----RTFMH-----GSNEGAVSILLR 64
 DB 1391 GNPSREFIVTGGPFRSTREERWBMQWESNRAIVLTR 1429

RESULT 10
 A49502

protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type 4E, splice form A precursor
 C:Species: Drosophila melanogaster
 C>Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 24-Apr-1998
 A:Accession: A49502

A:Molecule type: mRNA
A:Residues: 1-904,'L','V',906-1125,'Q','I',1126-1165,'YR','H',1168-1171,'A','I',1173-1215,'L','L',1217-1456,
A:Cross-references: GB:H80538; NID:g158644; PIDD:AAM28952.1; PID:g158645

C:Genetics:
A:Gene: FLYBase:Ptp10D
A:Cross-references: FLYBase:Fbgm0004370
C:Superfamily: protein-tyrosine-phosphatase, receptor type 4E; fibronectin type III repeat C;
C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; receptor
F:1197-1213/Domain: transmembrane #status predicted <TM>
F:1214-1557/Domain: intracellular #status predicted <INT>
F:1295-1515/Domain: protein-tyrosine-phosphatase homology <PTP>
F:1467/Active site: Cys (phosphocysteine intermediate) #status predicted
F:1473/Binding site: substrate phosphate (Arg) #status predicted

Dy Query Match 20.3%; Score 75; DB 2; Length 1557;
Best Local Similarity 26.8%; Pred. No. 1.3;
Matches 26; Conservative 10; Mismatches 19; Indels 42; Gaps 5;

OY 8 ELKLN-----PGFNNGNPSNRKNRYDLTLPQH-----GY 41
||||| | | | ||||| :
DB 1277 ELKHGRDPCTF--ADLCRNRPKNRFNTILPYDSRFKLQPVDDDEGSDYTINANYVGH 1334

OY 42 SGPNE-----RTFEWH---GSNEGAVSLLR 64
:-|- -||| | :-|
DB 1335 NSPREFIYTGCPPLHSIRDFWRMCWESNSRAIVMLTR 1371

RESULT 14.
CA1214
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type 10D, long splice form precured
C:Species: Drosophila melanogaster
C>Date: 28-May-1992 #sequence_revision 12-Jun-1992 #text_change 24-Apr-1998
C:Accession: CA1214
R:Xiang, X.; Seow, K.T.; Bahrl, S.M.; Oon, S.H.; Chia, W.
Cell 67, 661-673, 1991
A>Title: Two Drosophila receptor-like tyrosine phosphatase genes are expressed in a subset
A:Reference number: A41214; NCBI:2034988; PMID:1657401
A:Accession: CA1214
A:Molecule type: mRNA
A:Residues: 1-1630 <TM>
A:Cross-references: GB:H80465
C:Genetics:
A:Gene: FLYBase:Ptp10D
A:Cross-references: FLYBase:Fbgm0004370
C:Superfamily: protein-tyrosine-phosphatase, receptor type 4E; fibronectin type III repeat C;
C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; receptor
F:1197-1213/Domain: transmembrane #status predicted <TM>
F:1214-1630/Domain: intracellular #status predicted <INT>
F:1295-1515/Domain: protein-tyrosine-phosphatase homology <PTP>
F:1467/Active site: Cys (phosphocysteine intermediate) #status predicted
F:1473/Binding site: substrate phosphate (Arg) #status predicted

Dy Query Match 20.3%; Score 75; DB 2; Length 1630;
Best Local Similarity 26.8%; Pred. No. 1.4;
Matches 26; Conservative 10; Mismatches 19; Indels 42; Gaps 5;

OY 8 ELKLN-----PGFNNGNPSNRKNRYDLTLPQH-----GY 41
||||| | | | ||||| :
DB 1277 ELKHGRDPCTF--ADLCRNRPKNRFNTILPYDSRFKLQPVDDDEGSDYTINANYVGH 1334

OY 42 SGPNE-----RTFEWH---GSNEGAVSLLR 64
:-|- -||| | :-|
DB 1335 NSPREFIYTGCPPLHSIRDFWRMCWESNSRAIVMLTR 1371

A:Title: corkscrew encodes a putative protein tyrosine phosphatase that functions to
A:Reference number: A43254; MUID:92346711; PMID:1638629
A:Accession: A43254
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-841 <PER>
A:Cross-references: GB:M4730; NID:g157145 ; PID:g157145
A:Experimental source: embryo
A>Note: sequence extracted from NCBI backbone (NCBIN:109964, NCBIP:109965)
C:Genetics:
A:Gene: FlyBase:csw
A:Cross-references: FlyBase:FBgn0000382
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-phospha
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphina
F:6-101/Domain: SH2 homology <SH2>
F:111-203/Domain: SH2 homology <SH2B>
F:257-634/Domain: protein-tyrosine-phosphatase homology <PTP>
F:583/Active site: Cys (phosphocysteine intermediate) #status predicted
F:589/Binding site: substrate phosphate (Arg) #status predicted

	Query Match	19.8%	Score 73;	DB 2;	Length 841;
	Best Local Similarity	59.1%	Pred. No. 1.1;		
Matches	13; Conservative	3;	Mismatches	6;	Indels 0;
Gaps					0;

OY 17 NSGNQPSNRERNRRDILPEQH 38
| | : || |||||::|
Db 245 NEGKQENRLKNRYRNILPYDH 266

Search completed: June 6, 2003, 11:30:28
Job time : 16 secs

RESULT 15
A43254
protein-tyrosine-phosphatase (EC 3.1.3.46) corkscrew - fruit fly (*Drosophila melanogaster*)
C:Species: *Drosophila melanogaster*
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jun-2002
C:Accession: A43254
R:Perkins, L.A.; Larsen, I.; Perrimon, N.
Cell 70, 225-236, 1992

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 11:28:36 ; Search time 14 Seconds
(without alignments)
140.810 Million cell updates/sec

Title: US-09-935-703-7
Perfect score: 369
Sequence: 1 MOEFMALEKLNLPGEFNSGN.....TTFWHSGNVALSLRYCA 67

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88.5	24.0	699	1 US-08-348-006B-7	Sequence 7, Appl1
2	88.5	24.0	699	2 US-08-800-825A-7	Sequence 7, Appl1
3	88.5	24.0	699	4 US-09-158-657-7	Sequence 7, Appl1
4	88	23.8	513	1 US-08-202-389-2	Sequence 2, Appl1
5	86	23.3	595	1 US-08-202-389-6	Sequence 6, Appl1
6	86	23.3	631	1 US-08-202-389-8	Sequence 8, Appl1
7	79.5	21.5	793	1 US-08-015-985-3	Sequence 3, Appl1
8	79.5	21.5	802	1 US-08-015-985-3	Sequence 3, Appl1
9	78.5	21.3	155	4 US-08-821-278A-17	Sequence 17, Appl1
10	75.5	20.5	253	2 US-08-685-992-10	Sequence 10, Appl1
11	75.5	20.5	253	2 US-09-144-925-10	Sequence 10, Appl1
12	75	20.3	263	2 US-08-685-992-13	Sequence 13, Appl1
13	75	20.3	263	2 US-09-144-925-13	Sequence 13, Appl1
14	75	20.3	273	1 US-08-201-697-17	Sequence 17, Appl1
15	74	20.1	251	2 US-08-685-992-15	Sequence 15, Appl1
16	74	20.1	251	2 US-09-144-925-15	Sequence 15, Appl1
17	73.5	19.9	125	4 US-09-152-060-103	Sequence 103, App
18	73.5	19.9	471	4 US-09-152-060-102	Sequence 102, App
19	72	19.5	266	2 US-08-685-992-12	Sequence 12, Appl1
20	72	19.5	266	2 US-09-144-925-12	Sequence 12, Appl1
21	72	19.5	593	1 US-08-202-389-12	Sequence 12, Appl1
22	71	19.2	253	2 US-08-685-992-20	Sequence 20, Appl1
23	71	19.2	253	2 US-09-144-925-20	Sequence 20, Appl1
24	71	19.2	272	4 US-08-821-278A-19	Sequence 19, Appl1
25	71	19.2	802	4 US-09-081-345-18	Sequence 18, Appl1
26	70	19.0	593	1 US-08-018-129-5	Sequence 5, Appl1
27	70	19.0	593	2 US-08-448-250-5	Sequence 5, Appl1

28	69.5	18.8	254	2 US-08-685-992-9	Sequence 9, Appl1
29	69.5	18.8	254	4 US-09-144-925-9	Sequence 9, Appl1
30	69	18.7	278	4 US-08-821-278A-18	Sequence 18, Appl1
31	69	18.7	292	1 US-08-036-210-12	Sequence 12, Appl1
32	69	18.7	292	2 US-08-449-609-12	Sequence 12, Appl1
33	68	18.4	194	2 US-08-685-992-2	Sequence 2, Appl1
34	68	18.4	194	2 US-09-144-925-2	Sequence 2, Appl1
35	68	18.4	196	2 US-08-685-992-1	Sequence 1, Appl1
36	68	18.4	196	2 US-09-144-925-1	Sequence 1, Appl1
37	68	18.4	272	2 US-08-446-345-34	Sequence 34, Appl1
38	68	18.4	320	3 US-08-964-308-2	Sequence 2, Appl1
39	68	18.4	320	3 US-08-964-313-2	Sequence 2, Appl1
40	68	18.4	320	4 US-09-069-138-2	Sequence 2, Appl1
41	68	18.4	322	3 US-08-964-308-8	Sequence 8, Appl1
42	68	18.4	322	3 US-08-964-313-8	Sequence 8, Appl1
43	68	18.4	322	4 US-09-069-138-8	Sequence 8, Appl1
44	68	18.4	385	1 US-08-036-210-5	Sequence 5, Appl1
45	68	18.4	385	2 US-08-449-609-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-08-348-006B-7
Sequence 7, Application US/08348006B
Patent No. 5658756
GENERAL INFORMATION:
APPLICANT: RODAN, GIDEON A.
APPLICANT: SCHMIDT, AZRIEL
APPLICANT: RUTLEDGE, SU JANE
TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESSES:
ADDRESSEE: J. MARK HAND
STREET: 126 E. LINCOLN AVE., P.O. BOX 2000
CITY: RAHWAY
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/348,006B
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/122,032
FILING DATE: 14-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: HAND, J., MARK
REGISTRATION NUMBER: 36,545
REFERENCE/DOCKET NUMBER: 189921A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908-594-3905
TELEFAX: 908-594-4720
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 699 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-348-006B-7
Query Match 24.0%; Score 88.5; DB 1; Length 699;
Best Local Similarity 28.9%; Pred. No. 0.00061;
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,389
FILING DATE: 28-FEB-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/983,926
FILING DATE: 01-DEC-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/829,141
FILING DATE: 31-JAN-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/721,112
FILING DATE: 26-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BIH92-05MA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 513 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-202-389-2

Query Match 23.8%; Score 86; DB 1; Length 513;
Best Local Similarity 47.5%; Pred. No. 0.0063;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

QY 2 OEPMAL---ELKNLPGEFNSGNQPSNREKNRYRDLPEQH 38
:||:| |:||| |:||:||||:|||||
DB 164 EEFESLQKEVKNLHQRLEQGRPEKNKRNKRNILPEFH 202

RESULT 5
US-08-202-389-6
Sequence 6, Application US/08202389
Patent No. 5536636
GENERAL INFORMATION:
APPLICANT: Freeman Jr., Robert M.
APPLICANT: Plutsky, Jorge
APPLICANT: Neel, Benjamin G.
APPLICANT: Rosenberg, Robert D.
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE
NUMBER OF SEQUENCES: 54
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,389
FILING DATE: 28-FEB-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/983,926
FILING DATE: 01-DEC-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/829,141
FILING DATE: 31-JAN-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/721,112
FILING DATE: 26-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BIH92-05MA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/721,112
FILING DATE: 26-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BIH92-05MA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 595 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-202-389-6

Query Match 23.3%; Score 86; DB 1; Length 595;
Best Local Similarity 47.5%; Pred. No. 0.0015;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

QY 2 OEPMAL---ELKNLPGEFNSGNQPSNREKNRYRDLPEQH 38
:||:| |:||| |:||:||||:|||||
DB 246 EEFESLQKEVKNLHQRLEQGRPEKNKRNKRNILPEFH 284

RESULT 6
US-08-202-389-8
Sequence 8, Application US/08202389
Patent No. 5536636
GENERAL INFORMATION:
APPLICANT: Freeman Jr., Robert M.
APPLICANT: Plutsky, Jorge
APPLICANT: Neel, Benjamin G.
APPLICANT: Rosenberg, Robert D.
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE
NUMBER OF SEQUENCES: 54
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,389
FILING DATE: 28-FEB-1994
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/983,926
FILING DATE: 01-DEC-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/829,141
FILING DATE: 31-JAN-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/721,112
FILING DATE: 26-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BIH92-05MA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:

LENGTH: 631 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-202-389-8

Query Match
Best Local Similarity 23.3%; Score 86; DB 1; Length 631;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

QY 2 QEFMAL---ELKNIPEGFNSGNPSNRKRYRDLPPQH
DB 246 EEFESLQKDEKNIHQRL-EGQRENNKGRKRYKILPFDH 284

RESULT 7
US-08-015-985-3

Sequence 3, Application US/08015985
Patent No. 5538886
GENERAL INFORMATION:

APPLICANT: Schlessinger, Joseph
APPLICANT: Sap, Jan M.
TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/015,985
FILING DATE: 10-FEB-1993
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7683-020
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 793 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-015-985-3

Query Match
Best Local Similarity 21.5%; Score 79.5; DB 1; Length 793;
Matches 22; Conservative 9; Mismatches 19; Indels 33; Gaps 3;

QY 2 QEFMALEKNIPEGFNSGNPSNRKRYRDLPPQH
DB 234 EEFNALPACPIQATCEASKENKRNRYNILPDIHSRVHLTPVEGVPSDYINASFIN 293
QY 40 GY-----SGPNERTT--FW 51
DB 294 GYQENKFIAGQPKETVNDWF 316

RESULT 8
US-08-015-985-1
Sequence 1, Application US/08015985

Patent No. 5538886
GENERAL INFORMATION:

APPLICANT: Schlessinger, Joseph
APPLICANT: Sap, Jan M.
TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/015,985
FILING DATE: 10-FEB-1993
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7683-020
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 802 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-015-985-1

Query Match
Best Local Similarity 21.5%; Score 79.5; DB 1; Length 802;
Matches 22; Conservative 9; Mismatches 19; Indels 33; Gaps 3;

QY 2 QEFMALEKNIPEGFNSGNPSNRKRYRDLPPQH
DB 243 EEFNALPACPIQATCEASKENKRNRYNILPDIHSRVHLTPVEGVPSDYINASFIN 302
QY 40 GY-----SGPNERTT--FW 51
DB 303 GYQENKFIAGQPKETVNDWF 325

RESULT 9
US-08-821-278A-17

Sequence 17, Application US/08821278A
Patent No. 6238902
GENERAL INFORMATION:

APPLICANT: Cheng, Jili
APPLICANT: Lasky, Laurence A.
TITLE OF INVENTION: Protein Tyrosine Phosphatases
FILE REFERENCE: P1010R1
CURRENT APPLICATION NUMBER: US/08/821,278A
CURRENT FILING DATE: 1997-03-20
NUMBER OF SEQ ID NOS: 23
SEQ ID NO 17

LENGTH: 155
TYPE: PRT
ORGANISM: Homo Sapien

US-08-821-278A-17

Query Match
Best Local Similarity 21.3%; Score 78.5; DB 4; Length 155;
Matches 26; Conservative 10; Mismatches 19; Indels 51; Gaps 4;

Thu Jun 12 08:30:32 2003

us-09-935-703-7.rai

Page 5

QY	12	LPGEN-----	DSNPENREKNRYDILPE-----	QHNGYS-42
		: :	: :	
Db	10	LAGEFSIOACSAAMKADGVCSTVAGSRPFENVAKNNRYKDVLFPYDQTRVILSTLDEGSHD		69
QY	43	-----GPNERT--TFPHGNSNEGAVSLLRLRYC		66
		: :	: :	
Db	70	YINGNFIRGVDGSLAYIATGGLPRLHTLDDPWRLVMEGKAVILIMAC		115

RESULT 10
 US-08-685-992-10
 ? Sequence 10, Application US/08685992
 ? Patent No. 5912138
 ? GENERAL INFORMATION:
 ? APPLICANT: Tonks, Nicholas
 ? APPLICANT: Flint, Andrew J.
 ? TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
 ? TITLE OF INVENTION: TYROSINE PHOSPHATASES
 ? NUMBER OF SEQUENCES: 36
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
 ? STREET: Two Milltia Drive
 ? CITY: Lexington
 ? STATE: MA
 ? COUNTRY: USA
 ? ZIP: 02173
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Diskette
 ? COMPUTER: IBM Compatible
 ? OPERATING SYSTEM: Windows 95
 ? SOFTWARE: Fastseq for Windows Version 2.0b
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/685,992
 ? FILING DATE: 25-JUL-1996
 ? CLASSIFICATION: 435
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER:
 ? FILING DATE:
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Granahan, Patricia
 ? REGISTRATION NUMBER: 32,227
 ? REFERENCE/DOCKET NUMBER: CSHL96-03
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: 781-861-6240
 ? TELEFAX: 781-861-9540
 ? TELEX:
 ? INFORMATION FOR SEQ ID NO: 10:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 253 amino acids
 ? TYPE: amino acid
 ? STRANDEDNESS: single
 ? TOPOLOGY: linear
 ? MOLECULE TYPE: peptide
 ? US-08-685-992-10

```

Query Match          20.5%; Score 75.5; DB 2; Length 253;
Best Local Similarity 29.68; Pred. No. 0.014;
Matches      21; Conservative    3; Mismatches   14; Indels     33; Gaps       3

```



```

QY      14 GEENSGNOPSNRREKNRYRDILPFQH-----HGK-----S 42
        | | :| | | | | | | | | |
Db       2 GPEELANKNEERREKNRPNILPNDHSVILSQLDGIPCSXYINASYIDGYKEKKNFIAAO 61
        | | :| | | | | | | | | |
QY      43 GPNERTT--FW 51
        || :| | | | |
Db      62 GKPKETVNDWF 72

```



```

RESULT 11
US-09-144-925-10
; Sequence 10; Application US/09144925
; Patent No. 5951979

```

GENERAL INFORMATION:
APPLICANT: Tonks, Nicholas
TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
TITLE OF INVENTION: THIOSINE PHOSPHATASES
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
STREET: Two Millita Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02421-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/144,925
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/665,992
FILING DATE: July 25, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: CSHL96-03Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 781-861-6240
TELEFAX: 781-861-9540
TELEX:
INFORMATION FOR SRO ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-144-925-10

Query Match	20.5%	Score 75.5;	DB 2;	Length 253;
Best Local Similarity	29.6%;	Pred. No. 0.014;		
Matches	21;	Conservative	3;	Mismatches 14;
				Indels 33;
				Gaps 3

0Y 14 GEFNSGNOPSRNRRNRROLPLPFCH-----HG- S 42
 | : : ||||| : |||
 Db 2 GTFELANKEERREKRRYPNLPNDHSRYLLSQDLGPCSDYINASTYIDGKKENKFIAAG 61
 QY 43 GPNEKTT--FW 51
 | : : ||
 Db 62 GKQETVNDFW 72

RESULT 12
US-08-685-992-13
; Sequence 13, Application US/08685992
; Patent No. 5912138
; GENERAL INFORMATION:
; APPLICANT: Tonks, Nicholas
; APPLICANT: Flint, Andrew J.
; TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASES
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Millitta Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,992
FILING DATE: 25-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
TELECOMMUNICATION INFORMATION:
TELEPHONE: 781-861-6240
TELEFAX: 781-861-9540
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 263 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-685-992-13

Query Match
Best Local Similarity 20.3%; Score 75; DB 2; Length 263;
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 19 GNQPSNRKRNRYDILPFQ 38
DB 7 GOREPNKGNKRNKILPFQ 26

RESULT 13
US-09-144-925-13
Sequence 13, Application US/09144925
Patent No. 5951979
GENERAL INFORMATION:
APPLICANT: Tonks, Nicholas
TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
STREET: Two Millia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02421-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/144,925
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/685,992
FILING DATE: JULY 25, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
TELECOMMUNICATION INFORMATION:
TELEPHONE: 781-861-6240
TELEFAX: 781-861-9540

TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 263 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-144-925-13

Query Match
Best Local Similarity 20.3%; Score 75; DB 2; Length 263;
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 19 GNQPSNRKRNRYDILPFQ 38
DB 7 GOREPNKGNKRNKILPFQ 26

RESULT 14

US-08-201-697-17
Sequence 17, Application US/08201697
Patent No. 5705623
GENERAL INFORMATION:

APPLICANT: Wiggins, Roger C.
TITLE OF INVENTION: Mammalian Glomerular Epithelial Protein
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/201,697
FILING DATE: 25-FEB-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 273 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-201-697-17

Query Match
Best Local Similarity 20.3%; Score 75; DB 1; Length 273;
Matches 26; Conservative 10; Mismatches 19; Indels 42; Gaps 5;

OY 8 ELKNI-----PGEFNGQPSNRKRNRYDILPFQ 41
DB 28 ELKHVGRDPCF--ADLPCKRPNKRNRYDILPFQ 41
OY 42 SGPNE-----RTFEW--GSNEGAVSLR 64
DB 86 NSPREFVTOGLPHTSDRDFRMKCESNSRAIVMLR 122

RESULT 15

Thu Jun 12 08:30:32 2003

us-09-935-703-7.rapb

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2003, 11:30:11 ; Search time 45 Seconds
(without alignments)
153.713 Million cell updates/sec

Title: US-09-935-703-7
Perfect score: 1 MOEPMALELNKLPGEFNSGN.....TTFMHSGNEGAVSLIRYCA 67
Sequence: 1

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 392085 seqs, 103240269 residues

Total number of hits satisfying chosen parameters: 392085

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications-AA:*

- 1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
- 8: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
- 11: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	121	32.8	354	9	US-09-095-478-6
2	121	32.8	379	9	US-09-095-478-8
3	121	32.8	412	9	US-09-095-478-3
4	121	32.8	426	9	US-09-095-478-1
5	121	32.8	463	9	US-09-095-478-2
6	101	27.4	30	9	US-09-095-478-22
7	86	23.3	595	10	US-09-920-021A-1
8	86	23.3	843	9	US-10-072-036-117
9	86	23.3	853	9	US-10-072-036-119
10	78.5	21.3	458	9	US-10-087-993-36
11	75.5	20.5	309	10	US-09-788-626-11
12	75	20.3	322	10	US-09-788-626-14
13	74	20.1	129	10	US-09-788-626-16
14	73.5	19.9	125	10	US-09-852-797-103
15	73.5	19.9	125	10	US-09-853-161-103
16	73.5	19.9	125	10	US-09-852-659A-103
17	73.5	19.9	471	9	US-09-852-797-102
18	73.5	19.9	471	10	US-09-853-161-102
19	73.5	19.9	471	10	US-09-852-659A-102

20	72	19.5	325	10	US-09-788-626-13	Sequence 13, Appl
21	72	19.5	593	10	US-09-920-021A-3	Sequence 3, Appl
22	72	19.5	597	9	US-10-038-010-22	Sequence 22, Appl
23	71	19.2	312	10	US-09-788-626-21	Sequence 21, Appl
24	71	19.2	802	10	US-09-822-295-18	Sequence 18, Appl
25	70	19.0	180	10	US-09-925-300-1240	Sequence 1240, Ap
26	69.5	18.8	310	10	US-09-788-626-10	Sequence 10, Appl
27	68	18.4	251	10	US-09-788-626-3	Sequence 3, Appl
28	68	18.4	254	10	US-09-788-626-2	Sequence 2, Appl
29	68	18.4	435	10	US-09-986-240-1	Sequence 1, Appl
30	67.5	18.3	448	9	US-09-822-295-32	Sequence 32, Appl
31	67.5	18.3	453	9	US-10-087-993-32	Sequence 7, Appl
32	67	18.2	341	10	US-09-788-626-23	Sequence 23, Appl
33	67	17.9	264	9	US-10-245-539-6	Sequence 6, Appl
34	66	17.8	281	10	US-09-788-626-28	Sequence 28, Appl
35	64.5	17.5	281	10	US-09-788-626-3627	Sequence 3627, Ap
36	63.5	17.3	445	9	US-09-822-295-2	Sequence 2, Appl
37	62.5	16.9	1948	9	US-09-808-602-55	Sequence 55, Appl
38	62.5	16.9	1948	9	US-09-800-198-45	Sequence 45, Appl
39	62	16.8	316	10	US-09-788-626-5	Sequence 26, Appl
40	62	16.8	316	10	US-09-788-626-26	Sequence 4, Appl
41	62	16.8	317	10	US-09-788-626-4	Sequence 2, Appl
42	62	16.8	2308	12	US-10-000-954-9	Sequence 9, Appl
43	61	16.5	246	10	US-09-848-294-9	Sequence 28, Appl
44	61	16.5	250	10	US-09-848-294-8	Sequence 8, Appl
45	61	16.5	291	10	US-09-788-626-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-09-095-478-6
Sequence 6, Application US/09095478
Publication No. US20030095970A1
GENERAL INFORMATION:
APPLICANT: PLOMAN, Gregory
TITLE OF INVENTION: NOVEL PROTEIN TYROSINE
TITLE OF INVENTION: PHOSPHATASE SUPT05 AND
TITLE OF INVENTION: RELATED PRODUCTS AND
TITLE OF INVENTION: METHODS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FASTSEQ for Windows 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,478
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 224/115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FASTSEQ for Windows 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,478
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 224/115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 426 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Peptide
US-09-095-478-1

Query Match 32.8%; Score 121; DB 9; Length 426;
Best Local Similarity 61.1%; Pred. No. 1.4e-07;
Matches 22; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 1 MOEFMAELKNIPLGEFNSGNQPSNREKRYRDIPLF 36
Db 166 IREFLEQMTLPDPFNSGNTLNKRNKRYRDIPLF 201

RESULT 5
US-09-095-478-2
Sequence 2, Application US/09095478
Publication No. US20030095970A1
GENERAL INFORMATION:
APPLICANT: Plozman, Gregory
TITLE OF INVENTION: NOVEL PROTEIN TYROSINE
TITLE OF INVENTION: PHOSPHATASE SUPP05 AND
TITLE OF INVENTION: RELATED PRODUCTS AND
METHODS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FASTSEQ for Windows 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,478
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 224/115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Peptide
US-09-095-478-2

Query Match 32.8%; Score 121; DB 9; Length 463;
Best Local Similarity 61.1%; Pred. No. 1.6e-07;
Matches 22; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Qy 1 MOEFMAELKNIPLGEFNSGNQPSNREKRYRDIPLF 36
Db 203 IREFLEQMTLPDPFNSGNTLNKRNKRYRDIPLF 238

RESULT 6
US-09-095-478-22
Sequence 22, Application US/09095478
Publication No. US20030095970A1
GENERAL INFORMATION:
APPLICANT: Plozman, Gregory
TITLE OF INVENTION: NOVEL PROTEIN TYROSINE
TITLE OF INVENTION: PHOSPHATASE SUPP05 AND
TITLE OF INVENTION: RELATED PRODUCTS AND
METHODS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FASTSEQ for Windows 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,478
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 224/115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Peptide
US-09-935-478-22

Query Match

Best Local Similarity 27.4%; Score 101; DB 9; Length 30;
Matches 18; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 12 LPEFNSGNSQPSNREKRYRDLPP 36
DB 2 LPDPFNSGNTLQNRKRYRDLPP 26

RESULT 7

US-09-920-021A-1
Sequence 1, Application US/09920021A
Patent No. US20020110800A1
GENERAL INFORMATION:
APPLICANT: KAPLAN, DAVID
APPLICANT: MARSH, H. NICK
TITLE OF INVENTION: USE OF SHP-1 AND SHP-2 TO DETECT
FILE REFERENCE: 08338/016001
CURRENT APPLICATION NUMBER: US/09/920,021A
PRIOR FILING DATE: 2001-08-01
PRIOR APPLICATION NUMBER: US/08/918,157
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 595
TYPE: PRT
ORGANISM: Homo sapiens
US-09-920-021A-1

Query Match

Best Local Similarity 23.3%; Score 86; DB 10; Length 595;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

OY 2 QEFMAL---ELKNLPGEFNSGNSQPSNREKRYRDLPP 38
DB 246 EEFESLQKQEVKNLHQRL-EGORPENKGNKRYKNIILPFDH 284

RESULT 8

US-10-072-036-117
Sequence 117, Application US/10072036
Publication No. US20030082564A1
GENERAL INFORMATION:
APPLICANT: Ole THASTRUP
APPLICANT: Soren TULLIN
APPLICANT: Kasper ALMHOLT
APPLICANT: Kurt SCUDDER
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
FILE REFERENCE: 3759-0120P
CURRENT APPLICATION NUMBER: US/10/072,036
PRIOR FILING DATE: 2002-09-13
PRIOR APPLICATION NUMBER: 09/417,197
NUMBER OF SEQ ID NOS: 143
SOFTWARE: PatentIn version 3.0
SEQ ID NO 117
LENGTH: 843
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: EGFP-PTP fusion
US-10-072-036-117

Query Match 23.3%; Score 86; DB 9; Length 843;

Best Local Similarity 47.5%; Pred. No. 0.013;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

OY 2 QEFMAL---ELKNLPGEFNSGNSQPSNREKRYRDLPP 38

DB 494 EEFESLQKQEVKNLHQRL-EGORPENKGNKRYKNIILPFDH 532

RESULT 9

US-10-072-036-119
Sequence 119, Application US/10072036
Publication No. US20030082564A1
GENERAL INFORMATION:
APPLICANT: Ole THASTRUP
APPLICANT: Soren TULLIN
APPLICANT: Kasper ALMHOLT
APPLICANT: Kurt SCUDDER
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To A
FILE REFERENCE: 3759-0120P
CURRENT APPLICATION NUMBER: US/10/072,036
PRIOR FILING DATE: 2002-09-13
PRIOR APPLICATION NUMBER: 09/417,197
NUMBER OF SEQ ID NOS: 143
SOFTWARE: PatentIn version 3.0
SEQ ID NO 119
LENGTH: 853
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PTP-EGFP fusion
US-10-072-036-119

Query Match

Best Local Similarity 23.3%; Score 86; DB 9; Length 853;
Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

OY 2 QEFMAL---ELKNLPGEFNSGNSQPSNREKRYRDLPP 38

DB 248 EEFESLQKQEVKNLHQRL-EGORPENKGNKRYKNIILPFDH 286

RESULT 10

US-10-087-993-36
Sequence 36, Application US/10087993
Patent No. US20020169303A1
GENERAL INFORMATION:
APPLICANT: Ullrich, Axel

NOXI, Naohito
Kim, Yeong Woong
Wang, Hong Yang
Chen, Zhengjun
Naylor, Oliver
Kharitonov, Alexei Igorevich

TITLE OF INVENTION: NOVEL PTP20, PCP-2, BPL1, CLK,
AND SIRP POLYPEPTIDES AND RELATED
PRODUCTS AND METHODS

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSER: Lyon & Lyon
STREET: 633 West Fifth Street
Suite 4700

CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.

ZIP: 90071-2066

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 MB
storage

COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0


```

; Sequence 103, Application US/09852797
; Patent No. US20020172994A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003p2
; CURRENT APPLICATION NUMBER: US/09/852,797
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/265,583
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/152,060
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: PCT/US98/04858
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,357
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 103
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-852-797-103

```

```

Query Match      19.9%; Score 73.5; DB 9; Length 125;
Best Local Similarity 37.7%; Pred. No. 0.052;
Matches 20; Conservative 7; Mismatches 15; Indels 11; Gaps 2;

```

```

QY      9 LKNLGGEFNSGNQPSNRKRRYRDILPFQHHGYSGRN---ERTFWHGSNEG 57
||: | :||| | | | | | | | | | | | | | | | | | | | | | | | |
Db      28 LKSHAQERHSGNFPG-----RDSLPLFDQGHSGPPFANVEHSHFSYGARDG 73

```

```

RESULT 15
US-09-853-161-103
; Sequence 103, Application US/09853161
; Patent No. US20020076756A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: P2003p3
; CURRENT APPLICATION NUMBER: US/09/853,161
; CURRENT FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: 60/265,583
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/152,060
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: PCT/US98/04858
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/040,762
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/040,710
; PRIOR FILING DATE: 1997-03-14
; PRIOR APPLICATION NUMBER: 60/050,934
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,100
; PRIOR FILING DATE: 1997-05-30

```

```

; PRIOR APPLICATION NUMBER: 60/048,357
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/048,189
; PRIOR FILING DATE: 1997-05-30
; PRIOR APPLICATION NUMBER: 60/057,765
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/048,970
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 103
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-853-161-103

```

```

Query Match      19.9%; Score 73.5; DB 10; Length 125;
Best Local Similarity 37.7%; Pred. No. 0.052;
Matches 20; Conservative 7; Mismatches 15; Indels 11; Gaps 2;

```

```

QY      9 LKNLGGEFNSGNQPSNRKRRYRDILPFQHHGYSGRN---ERTFWHGSNEG 57
||: | :||| | | | | | | | | | | | | | | | | | | | | | | | |
Db      28 LKSHAQERHSGNFPG-----RDSLPLFDQGHSGPPFANVEHSHFSYGARDG 73

```

```

Search completed: June 6, 2003, 11:35:49
Job time : 46 secs

```

Thu Jun 12 08:30:31 2003

us-09-935-703-7.p2n.rst

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: June 11, 2003, 12:49:47 ; Search time 1435 Seconds

(without alignments)
756.166 Million cell updates/sec

Title: US-09-935-703-7

Perfect score: 369

Sequence: 1 MOEFMALEIKNLPGFNSGN.....TTFMHGNSNEGAVSILLRYCA 67

Scoring table: BLOSUM62
Xgapop 10.0, Ygapext 0.5
Ygapop 10.0, Ygapext 0.5
Xgapop 6.0, Ygapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xip
-Q/cgn2.1/uspro.spool/us09935703/runat_06062003_095211.6164/app.query.fasta_1.263
-DB=EST -QFMT=fastap -SUFFIX=pn.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=Blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09935703.ecgn.1.1.2874.grunat.06062003.095211.6164 -NCPU=6 -ICPU=3
-NO.MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=6 -YGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estlin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hic: *
9: gb_est1: *
10: gb_est2: *
11: gb_hic: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vit: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rtd: *

Pred. No. is the number of results predicted by cha...
score greater than or equal to the score of the result D...
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	263	71.3	571	13	BI460524
2	150	40.7	775	17	BG220714
3	145	39.3	719	17	AG752725
4	123	33.3	1842	11	AK017003
5	97.5	26.4	948	10	BM614565
6	88.5	24.0	604	10	AB221295
7	88.5	24.0	812	13	BI462537
8	88.5	24.0	1103	13	BM452119
9	88.5	24.0	2474	11	AK004602
10	88.5	24.0	2633	11	BC031078
11	88	23.8	369	13	BM429337
12	88	23.8	543	12	BF077448
13	88	23.8	727	13	BG962777
14	87.5	23.7	563	12	BE755886
15	87.5	23.7	1336	14	BM920649
16	86.5	23.4	999	13	BI768443
17	86	23.3	329	12	BE819448
18	86	23.3	349	10	AM408249
19	86	23.3	411	12	BF851252
20	86	23.3	413	10	AV696093
21	86	23.3	418	10	AM402786
22	86	23.3	436	12	BF762332
23	86	23.3	614	12	BF128847
24	86	23.3	818	12	BG398566
25	86	23.3	820	12	BG340572
26	86	23.3	849	12	BG398374
27	86	23.3	863	13	BI910569
28	86	23.3	865	14	BQ711049
29	86	23.3	919	14	BQ706936
30	84.5	22.9	347	14	D24553
31	84	22.2	339	12	BE831565
32	82	22.2	698	9	AJ395870
33	81.5	22.1	823	13	BM006699
34	81.5	22.1	943	13	BM006549
35	80.5	21.8	608	12	BG799244
36	80	21.5	514	10	AV398968
37	79.5	21.5	409	9	AA305783
38	79.5	21.5	531	9	AA572365
39	79.5	21.5	552	12	BG664544
40	79.5	21.5	563	10	BE303600
41	79.5	21.5	607	9	AU133714
42	79.5	21.5	613	10	AM412296
43	79.5	21.5	624	10	BE532150
44	79.5	21.5	639	14	BM787606
45	79.5	21.5	662	10	BE306537

ALIGNMENTS

RESULT 1
LOCUS BI460524 571 bp mRNA linear EST 21-AUG-2001
DEFINITION 603201139F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5267145 5',
mRNA sequence.
ACCESSION BI460524
VERSION BI460524.1 GI:15251180
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 571)
NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LLAM1673 row: n column: 10
High quality sequence stop: 569.

FEATURES
source

Location/Qualifiers
1..571
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site: 1: BamHI, Site: 2: SalI, XhoI (gtcag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTT-3',
size-selected for average insert size 2.2 kb and
normalized to 10⁶ 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI, National
Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT 172 a 117 c 150 g 132 t
ORIGIN

Alignment Scores:

Pred. No.: 3,27e-25 Length: 571
Score: 263.00 Matches: 48
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 71.27% Indels: 0
DB: 13 Gaps: 0

US-09-935-703-7 (1-67) x B1460524 (1-571)

QY 1 MetGInGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20
DB 427 ATGCAGAAATTTATGCTTACCTTAAAGAAATCTGCTGATCACTCTGGGAAT 486
QY 21 GlnProSerAsnArgGluLysAsnArgTyrArgAspLleuLeuProPheGlnHisGly 40
DB 487 CAACCAAGCAACAGAAAGAAAGATACCGAGATATCTTCATTCATCAATCAATGA 546
QY 41 TyrSerGlyProAsnGluArgThr 48
DB 547 TATAGTGGCCCAATGAGAGAACCA 570

RESULT 2
LOCUS BG220714 775 bp mRNA linear EST 21-APR-2001
DEFINITION R5740503 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG220714
VERSION BG220714.1 GI:13746735
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
Cain, S., Leventhal, C., Thornton, M., Ramachandran, P.D., Whittington, J.,
Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith,
J., Danzig, J., and Ducar, M.

TITLE Creation of genome-wide protein expression libraries using random
activation of gene expression

JOURNAL
MEDLINE
COMMENT

Nat. Biotechnol. 19 (5), 440-445 (2001)
21227151
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com
High quality sequence stop: 549.

FEATURES
source

Location/Qualifiers
1..775
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."

BASE COUNT 213 a 144 c 164 g 254 t
ORIGIN

Alignment Scores:

Pred. No.: 9.36e-10 Length: 775
Score: 150.00 Matches: 30
Percent Similarity: 90.91% Conservative: 0
Best Local Similarity: 90.91% Mismatches: 3
Query Match: 40.65% Indels: 1
DB: 12 Gaps: 0

US-09-935-703-7 (1-67) x BG220714 (1-775)

QY 35 PropheGlnHisGlyTyrSerGlyProAsnGluArgThrPheTrpHisGlySer 54
DB 1 CCTAGTCCATCATGAGATATGAGTGGCCCAATGAGAG-ACACGTTCTGGCATGTCTCA 59
QY 55 AsnGluGlyAlaValSerLeuLeuArgTyrCysAla 67
DB 60 AACGAGAGAGCAGATATCACTTTGTACCATATGTGCT 98

RESULT 3
LOCUS A0752725 719 bp DNA linear GSS 19-JUL-1999
DEFINITION HS_5568_B1_G06 RPCI-11 Human Male DNA Library Homo sapiens
ACCESSION A0752725
VERSION A0752725.1 GI:5539883
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Mahalas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D., and
Hood, L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
CONTACT: Mahalas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@jlong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)

FEATURES
source

Alignment Scores:	
Pred. No.:	3,96e-09
Score:	145.00
Percent Similarity:	96.67%
Best Local Similarity:	93.33%
Query Match:	39,30%
DB:	17
Length:	719
Matches:	28
Conservative:	1
Mismatches:	1
Indels:	0
Gaps:	0

US-09-935-703-7 (1-67) x AQ752725 (1-719)

6 AlaLeuGIuLeuLYAsnLeuProGIlyGIuPheAsnMeGIyAsnGIuProSerAsnArg 25
 495 GCCTTAGAACTTAGAATCTGCCTGCTGAGTTCAACTCTGGGATCAACCAAGCCACAGA 55

```

Oy      26 GLuLysAsnArgTyrArgAspIleLeuPro 35
          |||||
Db      555 GAAATAATAGATCCGAGATATTCTTCCA 584

```

LOCUS	AK017003	1842 bp	mRNA	linear	HTC 19-JAN-2002
DEFINITION	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:49343J18;protein tyrosine phosphatase, non-receptor type 20, full insert sequence.				

ACCESSION	AK017003
VERSION	AK017003.1
KEYWORDS	GI:12856041
SOURCE	HTC; CAP trapper.
	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA

clone:4933431F18.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus

REFERENCE
1
AUTHORS
TITLE
JOURNAL
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE	PUBMED	REFERENCE
20499374	11042159	
3		
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carinci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, K., Yamamoto, K., Matsumoto, H., Sakaiguchi, S., Ikegami, T., Kashiwagi, K.,		

MEDLINE
PUBMED

REFERENCE
AUTHORS

AUTHORS
Kawai, J., Shingawara, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
Akawa, T., Hara, A., Fukunishi, Y., Kono, H., Aadachi, J., Fukuda, S.,
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,
Saito, T., Okazaki, Y., Golobori, T., Bono, H., Kusunawa, T., Saito, R.,
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavert, T.,
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochwa, H.,
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Peesle, G.,
Quackenbush, J., Schriml, L. M., Stambli, F., Suzuki, R., Tomita, M.,
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
Cairncelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
Baldirelli, P., de Bonaldo, M. F., Brownstein, M. J., But, C.,
Pletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
Hormann, H., Hume, D. A., Kamata, M., Lee, N. H., Lyons, P.,
Marchionni, L., Mashima, J., Mazzaeili, J., Mommaerts, P., Nordone, P.,
Rang, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,
Toyo-oka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilmink, L.,
Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohguchi, S.,
and Hayashizaki, Y.

TITLE
Functional annotation of a full-length mouse cDNA collection
JOURNAL
Nature 409 (2001), 685-690 (2001)

PUBMED	11217851
REFERENCE	5 (bases 1 to 1842)
AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,

Akakawa, T., Baidarati, R., Bono, H., Brownstein, M., Bulc, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Horii, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, D., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishii, K., Nomura, K., Numasaki, R., Ono, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schirml, L., Shidata, K., Shidata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshinide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp, URL: <http://genome-gsc.riken.go.jp/>, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (<http://genome-gsc.riken.go.jp/>) for

For further details: cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGGACAGAGCATCCAGAGCTCTTTTCTTTTCTTTTAA 3'], cDNA was prepared by using triethanol thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot - 10.0 and subtraction to Rot - 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGGACAGATTCGAGTTAAATTATATCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised between 0.5 and 3 kb was selected before cloning. Vector: a modified pluscript KS(+) after bulk excision from lambda FLC I. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.

Alignment Scores:	
Pred. No.:	
Score:	1.51e-05
Percent Similarity:	123.00
Best Local Similarity:	66.00%
Query Match:	52.00%
DB:	33.33%
	11.33%
	Candels:
	1
	Length:
	1842
	Matches:
	26
	Conservative:
	7
	Mismatches:
	16
	Candels:
	1

1 MetGlnGluPheMetAlaLeuGlnIleuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20

2) GAGGCGCTGGTGGTTGGAGCTTAGAACAAATGACTCTGCCTGATGACCTTCATTCTGGGAAT 83

[illegible]

40 YTYRSerGlyProAsnGluArgThr 49
.....GCGGAGAAATTCCTCCATATGATTCACACCGT 89

897 GTTCCTCTGGAACCAAGACTACA 922

RESULT 5
BB614569

LOCUS	948 bp	mRNA	linear	EST 26-OCT-2003
DEFINITION	BB614569	BB614569	BB614569	BB614569
	RIKEN full-length enriched,	RIKEN full-length enriched,	RIKEN full-length enriched,	RIKEN full-length enriched,
	adult male testis cDNA	adult male testis cDNA	adult male testis cDNA	adult male testis cDNA

Accession	Gene	Accession	Gene
BB614569	4921513A12 5', mRNA sequence.	BB614569	4921513A12 5', mRNA sequence.
BB614569.1	GT-16A54050	BB614569.1	GT-16A54050

KEYWORDS EST.
SOURCE house mouse.
APPEARANCE

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelocephali; Mammalia; Eutheria; Euarchontoglires; Afrotheria; Rodentia; Muridae; Murinae; Murini; Mus

REFERENCE	AUTHORS	BASES	SEQUENCE
Arakawa, T., & Carninci, P.	Embryos, C.	1 (bases 1 to 948)	Scurognathi; Muridae; Mus.
Rodentia; Sciurognathi; Muridae; Murinae; Mus.			

Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, T., Okada, T., Furukawa, S., Furuno, M., Hangaki, T., Hara, A.

Ukazaki, Y., Okido, T., Salto, R., Sakai, C., Sakai, K., Sano, H.,
D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Teraami, M.

TITLE	Author	Year
RIKEN Mouse ESTs (Arikawa T et al)	Arikawa, T., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.	2000

Unpublished (2001)
Contact: Yoshihide Hayashizaki
Tokyo

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

Fax: 81-45-503-9216
 Email: genome-res@gsic.riken.go.jp,
 URL: <http://genome.gsic.riken.go.jp/>,
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh,
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. *Genome Res.* 10 (10), 1617-1630 (2000)
 Wajik, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura,
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. *Genome Res.*
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara,
 Y. and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)
 Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamakawa, I., Aizawa,
 K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
 Hayashizaki, Y.
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences. *Mamm. Genome* 12, 673-677 (2001)
 Please visit our web site (<http://genome.gsic.riken.go.jp>) for
 further details.
 e mouse tissues.
 Location/Qualifiers
 I. 948

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"

```

```

/clone="4921513A12"
/clone_lib="RIKEN full-length enriched, adult male testis"
/sex="male"

```

```

/issue_type="testis"
/dev_stage="adult"
/lab_host="SOLR"

```

/nucleo-site_1: XhoI; site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Center for Genome Dynamics, Iwatsubo, Japan.

Y. N. KAZAKH
 Biomedical Sciences Center and Genome Science Laboratory in
 RIKEN, Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with

GAGAGAGAGAGATCCACAGACTCTTTTITTTTTTTTNV [3], cDNA was prepared by using trehalase thermo-activated reverse transcription.

reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adaptor of sequence 15', GAGAGGAGATCCCCCCCCAATTA.

3']. CDNA was cloned into the XhoI and BamHI sites. "

Alignment Scores:	
0.0155	100%
red. No.:	0.0155

Score:	97.50	948
Local Similarity:	40.23%	30
Best Local Similarity:	34.48%	5
Misses:	0	0
Matches:	0	0
Length:	0	0

```

Query Match: 26.428
B:
Indels: 18
Gaps: 4

```

7 LeuclulLeulysasnLeuProGlgupheasnSerglvaancndBact-
S-09-935-703-7 (1-67) x BB614569 (1-948)

617 TTGAGACAAATGACTCTGCTGATGACTTCATTCCTTGCGGATACACTACAGAACAGACAT 676

*Lys*sna*r*gtyr*a*rgaspil*e*uProph*e*gl----- 37
|||||
677 AGAACAATACCGAGATTCTCCATGATGAACACCCTCCTCCCCCTTTC- 38

OY 38 -----HSHISGLY-----TyrSer 42
 DB 737 AGGATACATCAGCATATATAGATAGATACATGAGAGAGTATTTTATTCACACTCA 796
 OY 43 GlyPro-Aanglu-----ArgThrThrPheThrPhisGlySerAangluGlyAlaValse 60
 DB 797 GGGCCCTGCAGATATAGAGACTTTGGCAATGCTTCGGGNATPAT----- 842
 OY 60 fLeuLeuLeuArgTyrCys 66
 DB 843 -----TGTATCTATTGC 854
 RESULT 6
 AM321295 604 bp mRNA linear EST 25-JAN-2000
 LOCUS uc029g06.y1 NCI-CGAP Mam6 Mus musculus cDNA clone IMAGE:2616346 5'
 DEFINITION similar to gb:U35368 Mus musculus epsilon tyrosine phosphatase (MOUSE);, mRNA sequence.
 ACCESSION AM321295 GI:6750839
 VERSION AM321295.1
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 604)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 COMMENT Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 EMAIL: cgaaps-remail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: www-bio.lnl.gov/bdrp/image/image.html
 MGI:1024446
 Seq primer: -40RP from GIBCO
 High quality sequence stop: 433.
 Location/Qualifiers
 1..604
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:2616346"
 /clone_id="NCI-CGAP_Mam6"
 /sex="female, virgin"
 /tissue-type="infiltrating ductal carcinoma"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"
 BASE COUNT 177 a 147 c 158 g 122 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.13 Length: 604
 Score: 88.50 Matches: 24
 Percent Similarity: 37.35% Conservative: 7
 Best Local Similarity: 28.92% Mismatches: 19
 Query Match: 23.98% Indels: 33
 DB: 10 Gaps: 3
 US-09-935-703-7 (1-67) x AM321295 (1-604)
 OY 2 GlncluphneMetalaLeuGluLeuLysAnLeuProGlyGluPhenSerGlyAnGln 21

DB 44 GAGGAGTTCATTCCTTCGATCTGACACATACAGACACTTTGACACTGAC.
 OY 22 ProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis-----
 DB 104 GAAGAAAACAGAGAAAAAAGAGATACCCACATTCGCCAATGATCATTCAGAGTG 163
 OY 39 -----His 39
 DB 164 ATTTGAGCCAGTGTGATGAAATCCCTGCTGACATCATTAATGCTTCTACATCGAT 223
 OY 40 GlyTyr-----SerGlyProAsnGluArgThr---- 49
 DB 224 GGCTACAAAGAAAGAACAAATTCATACGACCTCAAGGCCCTTACAGAGACAGAT 283
 OY 50 --PheTyr 51
 DB 284 GACTCTCTGG 292
 RESULT 7
 B1462537 812 bp mRNA linear EST 21-AUG-2001
 LOCUS 603205447F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5271347 5',
 DEFINITION mRNA sequence.
 ACCESSION B1462537 GI:15253193
 VERSION B1462537.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 812)
 NIH-MGC http://mhc.ncbi.nlm.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 EMAIL: cgaaps-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shireki Toshlyuki and Piero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov
 Plate: LLM41684 row: m column: 12
 High quality sequence stop: 778.
 Location/Qualifiers
 1..812
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5271347"
 /clone_id="NIH_MGC_97"
 /lab_host="DH10B"
 /note="Organ: testis; Vector: pBluescriptPsr (modified pBluescript KS+); Site:1: BamHI; Site:2: SalI-XhoI (GCGAG 5'); Oligo-dT primed using primer 5'-TTTTTTTGTGTTT-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 234 a 200 c 224 g 154 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.202 Length: 812
 Score: 88.50 Matches: 24
 Percent Similarity: 37.35% Conservative: 7
 Best Local Similarity: 28.92% Mismatches: 19
 Query Match: 23.98% Indels: 33
 DB: 13 Gaps: 3

[illegible]

RESULT 8									
BM452119									
LOCUS									
DEFINITION	BM452119	1103 bp	mRNA	linear	EST 05-FEB-2007				
ACCESSION	AGENCE/COURT 6386102	NT_H_MCC_72	Homo sapiens	CDNA clone	IMAGE:5526462				
VERSION	BM452119								
KEYWORDS	BM452119.1	GI:18501159							
SOURCE	EST.								
ORGANISM	human.								
	Homo sapiens								

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominae; Homo.
1 (bases 1 to 1103)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC).
Unpublished (1999)
Contact: Robert Straussberg, Ph.D.

Email: c9apbs-remail.nih.gov
 Tissue Procurement: ATCC/DCMP
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LINTL)
 Clone distribution: MGC clone distribution information
 found through the I.M.A.G.E. Consortium/LINTL at:
<http://image.lnl.gov>
 plate: LLAM12200 row: C column: 07
 High quality sequence stop: 655.
 Location/Qualifiers
 1..1103

BASE COUNT	ORIGIN
294 a	322 c 274 g 212 t 1 others

```

Alignment Scores:
Pred. No.:      0.319
Score:          88.50
Percent Similarity: 37.35%
Best Local Similarity: 28.92%
Query Match:    23.98%
DB:             13

Length:         1103
Matches:        24
Conservative:   7
Mismatch:       19
Indels:         33
Gaps:           3

```

```

QY      2  GlnGluPheMetAlaLeuGluLeuLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln  21
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      286  GAGAGATTCAACTCATTTGCCATCTGGACACATACAGGAACCTTTTGAACTGGCAATATAA  344
QY      22  ProSerAsnArgGluLysAsnArgTyrTrpApIleLeuProPheGlnHis-----  38
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      346  GAGAGAAACAGAGAAAAAAACAGATATCCCAACATCTTCCCAATGACCATTTAGGGTG  402
QY      39  -----
Db      406  ATTCTGAGCCAACTGGATGGAATTCCCTGTTCAAGACTACATCATGCTTCTACATAGAT  465
QY      40  GlyTyr-----SerGlyProAsnGluArgThr---  49
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      466  GGTATACAAAGAGAAATAATTATAGACAGCTCAACGGTCCCAAAACAGAAAGGTTAAC  525
QY      50  ---PheTrp  51
      |||||
Db      526  GACTTCTTGG  534

```

RESULT 9	LOCUS	DEFINITION	AK004602
			AK004602
			Mus musculus adult male lung CDNA, RIKEN full-length enriched library, clone:1200004M14:protein tyrosine phosphatase, receptor type, E, full insert sequence.
			AK004602
			AK004602.1 GI:12835883
			HTC; CAP trapper.
			Mus musculus (strain:C57BL/6J) adult male

ORGANISM	REFERENCE
Mus musculus	Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
Mus musculus	
1	
2	
Carninci, P., Shidata, Y., Hayashizaki, Y., et al.	

TITLE
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 Genome Res. 10 (10), 1617-1630 (2000)
 JOURNAL
 MEDLINE
 20495374
 PUBMED
 11042159
 3
 REFERENCES
 AUTHORS
 Shibata, K., Itoh, M., Aizawa, K., Muramatsu, M., and Hayashizaki, Y.

TITLE	JOURNAL	GENOME RES.	DOI	PMID	PUBMED	REFERENCES
Kunjo, H., Akiyama, J., Nishik, K., Katsunai, T., Tasliro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, T., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.	RIKEN Integrated sequence analysis (RISA) system—384-format sequencing pipeline with 384 multicapillary sequencer	10 (11), 1757-1771 (2000)			11076861	

- 4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamakaki, I., Saito, T., Okazaki, Y., Gojohri, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Flieschmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuell, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staubli, F., Suzuki, R., Tomlita, M.,

```

/gene="p1pre-
/note="data source:MGI, source key:MGI:97813,
evidence:ISS
protein tyrosine phosphatase, receptor type, E"
/db_xref="MGI:MGI:97813"
BASE COUNT      627 a      659 c      692 g      436 t

```

Email: cgapbs-i@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shlirak
Toshyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94304
Web site: <http://www-simg.stanford.edu>
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MCC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://limage.lnl.gov>
Series: IRAT Plate: 48 Row: 1 Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA 91: 18860866
This clone has the following problem: frame shifted.
Location/Qualifiers
1..2633

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5295028"
/tissue_type="testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/notes="Vector: pBluescript"

BASE COUNT      702 a      694 c      681 g      556 t
ORIGIN

Alignment Scores:
Pred. No.:      1.17
Score:          88.50
Percent Similarity: 37.35%
Best Local Similarity: 28.92%
Query Match:    23.98%
                11
                Gaps: 3

US-09-935-703-7 (1-67) x BC031078 (1-2633)

QY      2 GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln 21
Db      690 GAGGAGTTCAACTGCTGGCCATCTGGACACATACAGGAACTTTTGAACCTGGCAATAAA 749
QY      22 ProSerAsnArgGlyLysAsnArgTyrArgAspIleLeuProPheGlnHis----- 38
Db      750 GAAGAAGAAAGAGAGAAAAAAGATATCCCAACATCTCCCATGACCATTTAGGGTG 809
QY      39 -----
Db      810 ATCTGAGCACTGATGATGATCCCTGTCAGACTACATCATGCTCTCTACATAGAT 869
QY      40 GlyTyr-----SerGlyProAsnGluAlyThr--- 49
Db      870 GGTACCAAGAGAGAAATATATCATAGACGCTCAAGGTCCCAAGCAAGAAAGGTTAAC 929
QY      50 ---PheTyr 51
Db      930 GACTCTCTGG 938

RESULT 11
LOCUS      BM429937      369 bp      mRNA      linear      EST 31-JAN-2002
DEFINITION 1Duo23D2.ab1 Bos taurus Duodenum #1 library Bos taurus CDNA, mRNA
ACCESSION  BM429937
VERSION    BM429937.1 GI:18451659
KEYWORDS   EST.
SOURCE     cow.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovidae; Bovine; Bos.
            1 (bases 1 to 369)
            Hansen,C., Fu,A., Meng,Y., Li,C., Okine,E., Senses,C.W., Gordon
            P.M.K. and Moore,S.S.
            Gene Expression Profiling of the Bovine Gastrointestinal Tract
            Unpublished (2002)
            Contact: Dr. Stephen Moore
            Beef Genomics Laboratory
            Dept of AFNS, University of Alberta
            410 Agri/Fer, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
            Tel: 780 492 0169
            Fax: 780 492 4265
            Email: smoores@afns.ualberta.ca
            Insert Length: 369 Std Error: 0.00
            POLY-A-No.

FEATURES
Source      Location/Qualifiers
            1..369
            /organism="Bos taurus"
            /db_xref="taxon:9913"
            /clone_lib="Bos taurus Duodenum #1 library"
            /tissue_type="Smooth muscle"

```

```

/cell_type="Simple columnar epithelial"
/dev_stage="Young adult"
/lab_host="XLI-BlueMR"strain"
/notes="Organ: Intestine/Duodenum; Vector: Uni-2ZAPXR;
       Site_1: ECORI; Site_2: Xho I"

BASE COUNT      105 a      94 c      110 g      60 t
ORIGIN

Alignment Scores:
Pred. No.:      0.0726
Score:          88.00
Percent Similarity: 65.00%
Best Local Similarity: 47.50%
Query Match:    23.85%
                13
                Gaps: 2

US-09-935-703-7 (1-67) x BM429937 (1-369)

QY      2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18
Db      129 GAGGAGTTTGAGAGTCTGCAGAGCAGAGAGTGAATAACTTGACACGCGGCTG---GAA 185
QY      19 GlyAsnGluProSerAsnArgGlyLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
Db      186 GGGCAGCGCGCCAGAAACAGAGACCAAGAACCGTTACAGAAACATCTCCGTTTGACAC 245

RESULT 12
LOCUS      BF077448      543 bp      mRNA      linear      EST 18-OCT-2000
DEFINITION 227352 MARC 2P1G Sus scrofa CDNA 5', mRNA sequence.
ACCESSION  BF077448
VERSION    BF077448.1 GI:10871278
KEYWORDS   EST.
SOURCE     pig.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
            1 (bases 1 to 543)
            Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keeler,J.W.
            Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
            Unpublished (2000)
            Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@maamr.ars.usda.gov
            Single pass sequencing. Bases called and alt trimmed with phred
            v0.980904.e. Vector identified by cross-match with the -mascote 18
            and -mismatch 12 options.
            PCR PRIMERS
            FORWARD: AGGAACAGCTATGACCAT
            BACKWARD: GTTTCACAGTCACGACG
            Plate: 51 row: F column: 7
            Seq primer: ATTAGGTGACACTATAG.

FEATURES
Source      Location/Qualifiers
            1..543
            /organism="Sus scrofa"
            /db_xref="taxon:9623"
            /clone_lib="MARC 2P1G"
            /tissue_type="pooled"
            /lab_host="DH10B"
            /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
            library made from pooled tissue from testis, ovary,
            endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT      144 a      142 c      170 g      87 t
ORIGIN

Alignment Scores:
Pred. No.:      0.129
Score:          88.00
Percent Similarity: 65.00%
Best Local Similarity: 47.50%
Query Match:    23.85%
                13
                Gaps: 2

```

Score: 88.00 Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.85% Indels: 4
DB: 12 Gaps: 2

US-09-935-703-7 (1-67) x BF077448 (1-543)

OY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18

DB 137 GAGGAGTTTCAAGAGCTGCGAAGACGACGAGGTGAAGAACTTGACACAGCGGCTG---GAA 193

OY 19 GlysAsnGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38

DB 194 GGGCAGCGCGCCAGAAACAGACAGACAGATCGTTACAGAAACATCTTCATTCGACACAC 253

RESULT 13 BG962777 727 bp mRNA linear EST 12-JUN-2001

LOCUS 602827607F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4982569 5',

DEFINITION mRNA sequence.

ACCESSION BG962777.1 GI:14350414

VERSION EST.

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Mus.

TITLE NIH-MGC http://mgc.nci.nih.gov/

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

EMAIL: cgaabs@remail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1M10985 row: m column: 02

High quality sequence stop: 704.

Location/Qualifiers

1..727

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone_1b="NCI_CGAP_Co24"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: Colon; Vector: pCMV-SPORT6; Site_1: NCI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP library."

BASE COUNT 202 a 182 c 208 g 135 t

ORIGIN

Alignment Scores:

Pred. No.: 0.2 Length: 727

Score: 88.00 Matches: 19

Percent Similarity: 65.00% Conservative: 7

Best Local Similarity: 47.50% Mismatches: 10

Query Match: 23.85% Indels: 4

DB: 13 Gaps: 2

US-09-935-703-7 (1-67) x BG962777 (1-727)

OY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18

DB 222 GAGGAGTTTCAAGAGCTGCGAAGACGACGAGGTGAAGAACTTGACACAGCGTCTG---GAA 278

OY 19 GlysAsnGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38

DB 279 GGGCAGCGCGCCAGAGAACAGACAAAGAACCCGCTACAGAACTTCTCCCTT.

RESULT 14

LOCUS BE755886

DEFINITION 209763 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION BE755886

VERSION BE755886.1 GI:10169815

KEYWORDS EST.

SOURCE COW.

ORGANISM Bos taurus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

TITLE Bovidae; Bovinae; Bos.

JOURNAL Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,

Caasas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,

G.L., Heaton,M.P., Laegreid,W.M., Rohrer,G.A., Chitko-McKown,C.G.,

Pertea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and

Keele,J.W.

Sequence evaluation of four pooled-tissue normalized bovine CDNA

libraries and construction of a gene index for cattle

Genome Res. 11 (4), 626-630 (2001)

2180013

CONTACT: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and alt-trimmed with phred

v0.980904.e. Vector identified by cross-match with the -minscore 18

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCATGACAT

BACKWARD: GTTTCACGACGACGAC

Plate: 57 row: I column: 22

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

1..563

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_1b="MARC 2BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI; Library made from pooled tissue from testis, thymus, semitendinosus muscle, longissimus muscle, pancreas, adrenal, and endometrium."

BASE COUNT 172 a 135 c 153 g 103 t

ORIGIN

Alignment Scores:

Pred. No.: 0.16 Length: 563

Score: 87.50 Matches: 24

Percent Similarity: 37.35% Conservative: 7

Best Local Similarity: 28.92% Mismatches: 19

Query Match: 23.71% Indels: 33

DB: 12 Gaps: 3

US-09-935-703-7 (1-67) x BE755886 (1-563)

OY 2 GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGln 21

DB 233 GAGGAGTTCAATCTCTGCGCTGACACAGTGAAGAACTTTGACATGGCCATATAA 292

OY 22 ProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis----- 38

DB 293 GAAGAAACAGAGAAAAAACAGATACCCCAACATCTCCCAATGACCAATTCAGGGGTG 352

OY 39 -----His 39

DB	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
DB	353	ATTTTAAAGCCAGTGCAGGAACCCCGGTTCAGACTACATCAACAGCTTCTACTATACAT	412									
OY	40	GLYTYR-----SerglyProAsnGluArgThr--	49									
DB	413	GGTTTAAAGAGAAAGATTAATTCATAGCTGCTCTCAAGGTCCCAAGCAGACAGTGAAT	472									
OY	50	--PheArg 51										
DB	473	GACTTCTGTG 481										
RESULT 15												
BM920649		1336 bp	mrna	linear	EST 12-MAR-2002							
BM920649												
AGENCOURT	6706025	NIH_MGC_115	Homo sapiens	cdna	clone	IMAGE:57518522						
BM920649												
BM920649.1		GI:19371028										
EST.												
human.												
Homo sapiens												
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;												
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.												
1 (bases 1 to 1336)												
NIH-MGC http://mgc.nci.nih.gov/.												
National Institutes of Health, Mammalian Gene Collection (MGC)												
Unpublished (1999)												
Contact: Robert Strausberg, Ph.D.												
Email: cgaabs-remail.nih.gov												
Tissue Procurement: Life Technologies, Inc.												
CDNA Library Preparation: Life Technologies, Inc.												
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)												
DNA Sequencing by: Agencourt Bioscience Corporation												
Clone distribution: MGC clone distribution information can be												
found through the I.M.A.G.E. Consortium/LNL at:												
http://image.lnl.gov												
Plate: LRAM12785 row: b column: 13												
High quality sequence start: 15												
High quality sequence stop: 344.												
Location/Qualifiers												
1. 1336												
/organism="Homo sapiens"												
/db_xref="taxon:9606"												
/clone="IMAGE:5751852"												
/clone_1lb="NIH_MGC_115"												
/lab_host="DH10B"												
/note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb. Insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."												
BASE COUNT	355	a	394	c	321	g	264	t	2	others		
ORIGIN												
Alignment Scores:												
Pred. No.:	0.581											
Score:	87.50											
Percent Similarity:	37.35%											
Best Local Similarity:	28.92%											
Query Match:	23.71%											
DB:	14											
US-09-935-703-7 (1-67) x	BM920649	(1-1336)										
OY	2	GlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerglyAsnGln	21									
Db	247	GAGGAGTTCAACATCATTCATCGACACATCAAGAACTTTGACATCGCAATTA	306									
OY	22	ProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis-----										

[illegible]

Search completed: June 11, 2003, 14:04:35
Job time : 1441 secs

XX WP1; 2001-565582/63.
 DR N-PSDB; AAH43389.
 XX
 PT Novel isolated protein tyrosine phosphatase polypeptide useful for
 PT treating and preventing Alzheimer's disease, rheumatoid arthritis,
 PT Grave's disease, arrhythmias, neoplasms, multiple sclerosis and
 PT diabetes mellitus
 XX
 PS Claim 1; Page 279-80; 286pp; English.
 XX
 CC The sequences given in AAB47496-500 show protein tyrosine phosphatase
 CC (PTPase) polypeptides. The PTPase polypeptides, or the cDNA encoding
 CC them, are useful for preventing, treating or ameliorating a medical
 CC condition in a mammalian subject. They are useful for diagnosing,
 CC preventing or treating neural disorders (e.g. Alzheimer's disease,
 CC Parkinson's disease), immune system disorders (e.g. Rheumatoid
 CC arthritis, Grave's disease), muscular disorders, reproductive
 CC disorders, pulmonary disorders, cardiovascular diseases (e.g. viral hepatitis),
 CC (e.g. arrhythmias), infectious diseases (e.g. neoplasms), diseases at the cellular
 CC level (e.g. multiple sclerosis), endocrine disorders (e.g. diabetes
 CC mellitus), and wound healing and epithelial cell proliferation.
 CC They are also useful for preventing hair loss, to prevent skin aging
 CC due to sunburn, to maintain organs before transplantation, to modulate
 CC mammalian characteristics, to treat weight disorders, to change a
 CC mammal's mental or physical state, or as a food additive or
 CC preservative.
 CC
 XX Sequence 67 AA:
 SO
 Query Match 100.0%; Score 369; DB 22; Length 67;
 Best Local Similarity 100.0%; Pred. No. 6.1e-46;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MOEFMAELKLNLPGEFNSGNOPSRNKKRYRDIILPFOHIGSGPNERFTFMHSGNAGVS 60
 DB 1 MOEFMAELKLNLPGEFNSGNOPSRNKKRYRDIILPFOHIGSGPNERFTFMHSGNAGVS 60
 QY 61 LLIRYCA 67
 DB 61 LLIRYCA 67
 QY 61 LLIRYCA 67
 DB 61 LLIRYCA 67
 RESULT 2
 ID AAE14454 standard; Protein; 420 AA.
 XX
 AC AAE14454:
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human protein phosphatase-4.
 XX
 KW Human; protein phosphatase; PP-4; immune system disorder; AIDS; allergy;
 KW neurological disorder; developmental disorder; Alzheimer's disease;
 KW cell proliferative disorder; Huntington's disease; arteriosclerosis;
 KW renal tubular acidosis; gonadal dysgenesis; cancer; adenocarcinoma;
 KW leukaemia; transgenic animal; gene therapy.
 XX
 OS Homo sapiens.
 XX
 FT Key Location/Qualifiers
 FT Region 183..411
 FT /note="Protein-tyrosine phosphatase"
 FT Active-site 340..388
 FT /note="Tyrosine specific protein phosphatase active
 FT site"
 FT Domain 362..379
 FT /label="Transmembrane_domain
 FT Region 351..363
 FT /note="Tyr-phosphatase"
 XX

PN MO200196546-A2.
 XX
 PD 20-DEC-2001.
 XX
 PF 14-JUN-2001; 2001MO-US19442.
 XX
 PR 16-JUN-2000; 2000US-212447P.
 PR 22-JUN-2000; 2000US-213746P.
 PR 29-JUN-2000; 2000US-215210P.
 PR 06-JUL-2000; 2000US-216529P.
 PR 12-JUL-2000; 2000US-218080P.
 PR 21-JUL-2000; 2000US-220117P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Au-Yang J, Baughn MR, Ding L, Elliott VS, Gandhi AR, Griffin JA;
 PI Hafalla A, Kearney L, Lee EA, Lu Y, Nguyen DB, Patterson C;
 PI Ramkumar J, Reddy R, Sanjanwala MS, Stewart EA, Tang YF;
 PI Thornton M, Tribouley CM, Walla NK, Yang J, Yao MG, Yue H;
 XX
 DR WP1; 2002-090206/12.
 DR N-PSDB; AAD24022.
 XX
 PT Novel polypeptide, useful for diagnosing, treating or preventing
 PT disorders of growth and development, immune system, neurological and
 PT cell proliferation diseases, comprises cancer protein phosphatase
 PT polypeptides
 XX
 PS Claim 1; Page 105-106; 116pp; English.
 XX
 CC The present sequence is human protein phosphatase (PP)-4.
 CC PP polynucleotide and polypeptide are useful in the diagnosis,
 CC treatment and prevention of immune system disorders, neurological
 CC disorders, developmental disorders and cell proliferative disorders.
 CC Examples of immune system disorders include acquired immune deficiency
 CC syndrome (AIDS), severe combined immunodeficiency disease (SCID),
 CC adult respiratory distress syndrome, allergies, amyloidosis,
 CC anaemia, asthma, atherosclerosis, Crohn's disease, atopic
 CC dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,
 CC Graves' disease, multiple sclerosis, myasthenia gravis, myocardial or
 CC pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis,
 CC psoriasis, Reiter's syndrome, rheumatoid arthritis, Sjogren's syndrome,
 CC scleroderma, systemic sclerosis, trauma; neurological disorders include
 CC Alzheimer's disease, Huntington's disease, dementia, epilepsy,
 CC Parkinson's disease, mental retardation and other developmental
 CC disorders of central nervous system such as Down's syndrome, cerebral
 CC palsy, periodic paralysis, mental disorders including mood, anxiety,
 CC and schizophrenic disorders, seasonal affective disorder such as
 CC akathesia, amnesia, catatonia, dyskinesia; developmental disorders
 CC include e.g. renal tubular acidosis, Duchenne and Becker muscular
 CC dystrophy, gonadal dysgenesis, hypothyroidism; cell proliferative
 CC disorders include e.g. actinic keratosis, arteriosclerosis,
 CC atherosclerosis, bursitis, cirrhosis, hepatitis, psoriasis, and
 CC cancer including adenocarcinoma, leukaemia. The polypeptide and
 CC polynucleotide are further useful for analysing proteins of a tissue
 CC or a cell type, for screening an agonist/antagonist, a compound that
 CC specifically binds to it or its modulator. The polynucleotide is useful
 CC for creating knockin humanised animals (pigs) or transgenic animals
 CC (mice or rats) to model human disease, for generating a transcript image
 CC of a tissue or cell type, which represents the global pattern of gene
 CC expression by a particular tissue or cell type.
 CC
 XX Sequence 420 AA:
 SO
 Query Match 50.9%; Score 188; DB 23; Length 420;
 Best Local Similarity 97.2%; Pred. No. 1.8e-18;
 Matches 35; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MOEFMAELKLNLPGEFNSGNOPSRNKKRYRDIILP 36
 DB 160 MOEFMAELKLNLPGEFNSGNOPSRNKKRYRDIILP 195

RESULT 3

ABG06042

ID ABG06042 standard; Protein; 561 AA.

AC ABG06042;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #6033.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS70229.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT biobiodiversity for genetic disorders or other traits and to assess
PT biobiodiversity

XX Claim 20; SEQ ID NO 36401; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridization probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biobiodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 561 AA;

Query Match 48.8%; Score 180; DB 22; Length 561;

Best Local Similarity 94.4%; Pred. No. 3.8e-17;

Matches 34; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MOEFMALEKLNLPGEFNSGNOPSNRKRNRYRDIILPF 36

DB 301 MOEFMALEKLNLPGEFNSGNOPSNRKRNRYRDIILPF 336

RESULT 4

AAM89251

ID AAM89251 standard; Protein; 405 AA.

AC AAM89251;

DT 10-MAR-1999 (first entry)

DE Mouse PTP05 isoform #2.

KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease.

OS Mus sp.

XX WO9849317-A2.

XX 05-NOV-1998.

XX 27-APR-1998; 98WO-US08439.

XX 23-OCT-1997; 97US-0063595.

XX 28-APR-1997; 97US-0044428.

XX 20-MAY-1997; 97US-0047222.

XX 11-JUN-1997; 97US-0049477.

XX 18-JUN-1997; 97US-0049756.

XX 18-JUN-1997; 97US-0049914.

XX (SUGC-) SUGEN INC.

XX App H, Clary D, Courtneidge SA, Hui TH, Jallat B;

XX Marky D, Onrust S, Peles E, Plozman GD;

XX WPI; 1999-009434/01.

XX N-PSDB; AAV81746.

PT New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease

XX Claim 2; Page 158-160; 193pp; English.

XX The present invention describes isolated, enriched or purified nucleic
XX acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
XX present sequence represents mouse PTP05. The above proteins, other than
XX ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
XX substances that modulate their activity (i.e. agonists and antagonists,
XX including NBP) in vivo or in vitro. These substances are used to treat
XX or prevent diseases associated with abnormal signal transduction
XX pathways that involve the proteins, particularly cancer (e.g. leukaemia
XX and lymphoma), while modulators of ALK-7 (which is a type I receptor
XX serine/threonine kinase) are used to promote neuronal survival.
XX Particularly for treating Alzheimer's, Parkinson's or Huntington's
XX diseases. Nucleic acid fragments of the polynucleotides encoding the
XX proteins can be used as probes to identify and clone related sequences;
XX to detect protein-encoded RNA; to generate transgenic animals and in
XX gene therapy (optionally after mutation). Ab are used to determine the
XX proteins.

SQ Sequence 405 AA;

Query Match 32.8%; Score 121; DB 20; Length 405;

Best Local Similarity 61.1%; Pred. No. 1e-08;

Matches 22; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 MOEFMALEKLNLPGEFNSGNOPSNRKRNRYRDIILPF 36

DB 166 IREFLEQMTLPDDFNSGNLTIONDRDKNRYRDIILPF 201

RESULT 5

AAM89249

ID AAM89249 standard; Protein; 426 AA.

```

XX AC AAW89249;
XX DT 10-MAR-1999 (first entry)
XX DE Mouse PTP05.
XX KW PPP04; PPP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;
XX KW type I receptor serine/threonine kinase; cancer; leukemia; lymphoma;
XX KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
XX KW Parkinson's disease; Huntington's disease.
XX OS Mus sp.
XX PN W09849317-A2.
XX PD 05-NOV-1998.
XX PF 27-APR-1998; 98WO-US08439.
XX PR 23-OCT-1997; 97US-0063595.
XX PR 28-APR-1997; 97US-0044428.
XX PR 20-MAY-1997; 97US-0047222.
XX PR 11-JUN-1997; 97US-0049477.
XX PR 11-JUN-1997; 97US-0049756.
XX PR 18-JUN-1997; 97US-0049914.
XX PA (SUGEN-) SUGEN INC.
XX PI App H, Clary D, Courtneldge SA, Hui TH, Jallal B;
XX PI Markby D, Onrust S, Peles E, Plovman GD;
XX DR N-PSDB; AAW81744.
XX DR WPI; 1999-009434/01.
XX PT New nucleic acid encoding specific protein tyrosine phosphatases -
XX PT useful for identifying specific modulators for treatment and
XX PT prevention of cancer and neurodegenerative disease
XX PS Claim 2; Page 155-157; 193pp; English.
XX CC The present invention describes isolated, enriched or purified nucleic
XX CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
XX CC present sequence represents mouse PTP05. The above proteins, other than
XX CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
XX CC substances that modulate their activity (i.e. agonists and antagonists,
XX CC including NBP) in vivo or in vitro. These substances are used to treat
XX CC or prevent diseases associated with abnormal signal transduction
XX CC pathways that involve the proteins, particularly cancer (e.g. leukemia
XX CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
XX CC serine/threonine kinase) are used to promote neuronal survival,
XX CC particularly for treating Alzheimer's, Parkinson's or Huntington's
XX CC diseases. Nucleic acid fragments of the polynucleotides encoding the
XX CC proteins can be used as probes to identify and clone related sequences;
XX CC to detect protein-encoded RNA; to generate transgenic animals and in
XX CC gene therapy (optionally after mutation). Ab are used to determine the
XX CC proteins.
XX SQ Sequence 426 AA;
XX QY Query Match 32.8%; Score 121; DB 20; Length 426;
XX QY Best Local Similarity 61.1%; Pred. No. 1.1e-08;
XX QY Matches 22; Conservative 6; Mismatches 8; Indels 0; Gaps 0;
XX DB 1 MOEFMALEKLNLPGEFNSGNOPSNRKRYRDLIPF 36
XX DB 166 IREFLEQMTLPDDFNSGNTLQNRDKRYRDLIPY 201

```

RESULT 6
AAW89250
ID AAW89250 standard; Protein; 463 AA.
XX

```

AC AAW89250;
XX DT 10-MAR-1999 (first entry)
XX DE Mouse PTP05 isoform #1.
XX KW PPP04; PPP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;
XX KW type I receptor serine/threonine kinase; cancer; leukemia; lymphoma;
XX KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
XX KW Parkinson's disease; Huntington's disease.
XX OS Mus sp.
XX PN W09849317-A2.
XX PD 05-NOV-1998.
XX PF 27-APR-1998; 98WO-US08439.
XX PR 23-OCT-1997; 97US-0063595.
XX PR 28-APR-1997; 97US-0044428.
XX PR 20-MAY-1997; 97US-0047222.
XX PR 11-JUN-1997; 97US-0049477.
XX PR 11-JUN-1997; 97US-0049756.
XX PR 18-JUN-1997; 97US-0049914.
XX PA (SUGEN-) SUGEN INC.
XX PI App H, Clary D, Courtneldge SA, Hui TH, Jallal B;
XX PI Markby D, Onrust S, Peles E, Plovman GD;
XX DR N-PSDB; AAW81745.
XX DR WPI; 1999-009434/01.
XX PT New nucleic acid encoding specific protein tyrosine phosphatases -
XX PT useful for identifying specific modulators for treatment and
XX PT prevention of cancer and neurodegenerative disease
XX PS Claim 2; Page 157-158; 193pp; English.
XX CC The present invention describes isolated, enriched or purified nucleic
XX CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
XX CC present sequence represents mouse PTP05. The above proteins, other than
XX CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
XX CC substances that modulate their activity (i.e. agonists and antagonists,
XX CC including NBP) in vivo or in vitro. These substances are used to treat
XX CC or prevent diseases associated with abnormal signal transduction
XX CC pathways that involve the proteins, particularly cancer (e.g. leukemia
XX CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
XX CC serine/threonine kinase) are used to promote neuronal survival,
XX CC particularly for treating Alzheimer's, Parkinson's or Huntington's
XX CC diseases. Nucleic acid fragments of the polynucleotides encoding the
XX CC proteins can be used as probes to identify and clone related sequences;
XX CC to detect protein-encoded RNA; to generate transgenic animals and in
XX CC gene therapy (optionally after mutation). Ab are used to determine the
XX CC proteins.
XX SQ Sequence 463 AA;
XX QY Query Match 32.8%; Score 121; DB 20; Length 463;
XX QY Best Local Similarity 61.1%; Pred. No. 1.2e-08;
XX QY Matches 22; Conservative 6; Mismatches 8; Indels 0; Gaps 0;
XX DB 1 MOEFMALEKLNLPGEFNSGNOPSNRKRYRDLIPF 36
XX DB 203 IREFLEQMTLPDDFNSGNTLQNRDKRYRDLIPY 238

```

RESULT 7
AAW27226
ID AAW27226 standard; Protein; 699 AA.
XX
AC AAW27226;


```

XX Human: sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSEQ-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR MPI: 2001-476164/51.
DR N-PSDB: AAH98486.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 20; Page 953-954; 1275pp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
SQ Sequence 700 AA;
Query Match 24.0%; Score 88.5; DB 22; Length 700;
Best Local Similarity 28.9%; Pred. No. 0.0011;
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;
QY 2 OEFALELKLNPGEFNSGNQPSNREKNRYRDLPFQH-----H 39
DB 137 EEFNLSLPSGHIOGTFFELANKRENEKNRYPNILPDHSRVILSOLDGIPCSDIYINASTID 196
QY 40 GY-----SGPNERTT--FW 51
DB 197 GYKEKNKFTIAAGPKQETVNDFW 219

```

RESULT 12
ID AAM38831 standard; Protein: 700 AA.
AC AAM38831;
XX
XX 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 1976.
XX
KW Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX

```

OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0633450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSEQ-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR MPI: 2001-442253/47.
DR N-PSDB: AAI57987.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 3; SEQ ID NO 1976; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AA42213) with noctropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localized neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC activation/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 700 AA;
Query Match 24.0%; Score 88.5; DB 22; Length 700;
Best Local Similarity 28.9%; Pred. No. 0.0011;
Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;
QY 2 OEFALELKLNPGEFNSGNQPSNREKNRYRDLPFQH-----H 39
DB 137 EEFNLSLPSGHIOGTFFELANKRENEKNRYPNILPDHSRVILSOLDGIPCSDIYINASTID 196
QY 40 GY-----SGPNERTT--FW 51
DB 197 GYKEKNKFTIAAGPKQETVNDFW 219

```

RESULT 13
ID AAM40617 standard; Protein: 717 AA.
AC AAM40617;
XX
XX 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 5548.
XX

Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokine; thrombolytic; drug screening; arthritis; inflammation;
leukemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0634350.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QH, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX N-PSDB; AA159773.

XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -

XX Example 2; SEQ ID NO 5548; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AA138642-AA142213) with noctropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localized neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukemias and
XX C.N.S disorders.

XX Note: The sequence data for this patent did not form part of the printed
XX specification.

XX Sequence 717 AA;

XX Query Match 24.0%; Score 88.5; DB 22; Length 717;
XX Best Local Similarity 28.9%; Pred. No. 0.0012;
XX Matches 24; Conservative 7; Mismatches 19; Indels 33; Gaps 3;

XX 2 QEFMALEKNTPEGNSGNOPSNRKKNRYRDLTPFH-----H 39

XX 154 EEFNSLPSGHIGTFELANKNEKKNRPNLPNDHSVILLQDLGIPCSDYINASYID 213

XX 40 GY-----SGPNERTT--FW 51

XX 214 GYKKNKFIAMQGPKEFTVNDWF 236

XX RESULT 14

XX AAR99314

XX ID AAR99314 standard; Protein; 513 AA.

XX AAR99314;

XX 13-NOV-1996 (first entry)

XX Rat M1PRP (protein tyrosine phosphatase-1) clone.

XX PTP; protein tyrosine phosphatase; SH2; Src homology region 2;
XX chromosome 12p; abnormality; mutation; detection; probe; neoplasia;
XX cancer; leukemia; diagnosis; megakaryocyte regulation.

XX Rattus sp.

XX US5536636-A.

XX 16-JUL-1996.

XX 26-JUN-1991; 91US-0721112.

XX 28-FEB-1994; 94US-0202389.

XX 26-JUN-1991; 91US-0721112.

XX 31-JAN-1992; 92US-0829141.

XX 01-DEC-1992; 92US-0983926.

XX (BETH-) BETH ISRAEL HOSPITAL ASSOC.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX Freeman RM, Neel BG, Plutsky J, Rosenberg RD;

XX WPI; 1996-341506/34.

XX N-PSDB; AAR35308.

XX Detecting 12p chromosomal abnormality associated with neoplastic
XX disease - using SH-PRP1 protein tyrosine phosphatase gene specific
XX probe

XX Example 1; Column 32-34; 63pp; English.

XX AAR99314 is a rat M1PRP (protein tyrosine phosphatase-1) clone. A cDNA
XX clone encoding this sequence was used as a probe to isolate a
XX series of overlapping human cDNA clones defining the human M1PRP
XX equivalent, SH-PRP1, which contains two tandem SH2 domains. A
XX sequence complementary to nucleotides 537-653 of the SH-PRP1 gene
XX clone given as AAR35306 is used as a probe to detect a chromosome
XX 12p13 abnormality associated with neoplastic disease, in partic.
XX acute lymphoblastic leukemia. The probe hybridises to a part
XX of the region coding for the two tandem SH2 domains (see AAR99312).

XX If the probe will not hybridise DNA of chromosome 12p13 from a patient
XX sample it is indicative of an abnormality, normally associated with
XX neoplasia. Fragments of sequences encoding human SH-PRP2 (see AAR35307)
XX may also be used to diagnose a condition or susceptibility to a
XX condition associated with chromosome 12p abnormality. Alternatively
XX the wild-type SH-PRP1 or SH-PRP2 gene or protein may be used for
XX comparison to sequenced PTP genes taken from a patient, where
XX differences indicate an abnormality. The activity of SH-PRP1 may also
XX be purposely altered by mutation to effect a change in megakaryocyte
XX function and hence platelet production.

XX Sequence 513 AA;

XX Query Match 23.8%; Score 88; DB 17; Length 513;
XX Best Local Similarity 47.5%; Pred. No. 0.00089;
XX Matches 19; Conservative 7; Mismatches 10; Indels 4; Gaps 2;

XX 2 QEFMAL---ELKMLPGEFNSGNOPSNRKKNRYRDLTPFH 38

XX 164 EEFESLQKQEVKNLHQRLL-EGORPENKSKKNRYNLPFDH 202

XX RESULT 15

XX AAR99316

XX ID AAR99316 standard; Protein; 458 AA.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 11, 2003, 12:55:32 ; Search time 64 Seconds
(Without alignments)
321,052 Million cell updates/sec

Title: US-09-935-703-7

Perfect score: 369
Sequence: 1 MOEFMALEIKLKPGEFNSGN.....TTFMGHSGNAGVSLIRYCA 67

Scoring table:

BLOSUM62
Xgapop 10.0, Ygapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL-frame+P2n.model -DEV-xlp
-Q=/cgn2_1/USPRO.spool/US09935703/runat_06062003_095212.6181/app-query.fasta.1.263
-DB-issued_patents_NA -QEMT-fastcap -SUFFIX-p2n.rml -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE-pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE-LOCAL -OUTFMT=pro -NORR-ext HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09935703.ecgn.1.1.61.etruncat.06062003_095212.6181 -NCP=6 -ICPU=3
-NO_MMAP -LARGESOURCY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV-TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database:

1: Issued_Patents_NA:*
2: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/BACKFILES1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88	23.8	1747	1	US-08-202-389-1
2	88	23.3	2143	1	US-08-202-389-7
3	86	23.3	2145	1	US-08-202-388-5
4	86	23.3	2277	3	US-09-358-685-1
5	79.5	21.5	2409	1	US-08-015-985-2
6	79.5	21.5	2872	1	US-08-015-985-4
7	78.5	21.3	466	4	US-08-821-2278A-16
8	73.5	19.9	1659	4	US-09-152-060-24
9	73	19.8	878	2	US-08-446-345-33
10	73	19.8	4080	2	US-08-446-345-35
11	72	19.5	2121	4	US-09-358-683-1
12	72	19.5	2276	1	US-08-202-389-11

13	71	19.2	247	1	US-08-594-031-74	Sequence 74, Appl
14	70	19.0	612	4	US-09-280-116-218	Sequence 218, App
15	70	19.0	2790	2	US-08-018-129-4	Sequence 4, Appl
16	70	19.0	2790	2	US-08-448-250-4	Sequence 4, Appl
17	68	18.4	963	3	US-08-964-308-1	Sequence 1, Appl
18	68	18.4	963	3	US-08-964-308-7	Sequence 7, Appl
19	68	18.4	963	3	US-08-964-313-1	Sequence 1, Appl
20	68	18.4	963	3	US-08-964-313-7	Sequence 7, Appl
21	68	18.4	963	4	US-09-069-138-1	Sequence 1, Appl
22	68	18.4	963	4	US-09-069-138-7	Sequence 7, Appl
23	68	18.4	3247	4	US-09-487-368A-3	Sequence 3, Appl
24	68	18.4	3969	1	US-08-241-853-1	Sequence 1, Appl
25	68	18.4	4369	2	US-08-850-917-1	Sequence 1, Appl
26	68	18.4	4127	4	US-09-487-368A-10	Sequence 10, Appl
27	67.5	18.3	2226	3	US-08-951-260A-1	Sequence 1, Appl
28	67.5	18.3	4403765	4	US-09-103-840A-2	Sequence 2, Appl
29	67.5	18.3	4411529	4	US-09-103-840A-1	Sequence 1, Appl
30	66.5	18.0	4752	1	US-08-201-697-3	Sequence 3, Appl
31	66.5	18.0	4815	1	US-08-201-697-5	Sequence 5, Appl
32	66.5	18.0	5679	1	US-08-201-697-1	Sequence 1, Appl
33	65.5	17.8	4799	1	US-08-201-697-6	Sequence 6, Appl
34	64.5	17.5	3580	4	US-09-081-345-1	Sequence 1, Appl
35	63	17.1	1079	4	US-09-221-017B-540	Sequence 540, App
36	62.5	16.9	2577	1	US-08-439-818A-4	Sequence 4, Appl
37	62.5	16.9	2577	2	US-08-751-965-4	Sequence 4, Appl
38	62.5	16.9	2577	2	US-08-738-975-4	Sequence 4, Appl
39	62.5	16.9	2577	2	US-08-728-626-4	Sequence 4, Appl
40	62.5	16.9	2577	3	US-08-808-599A-4	Sequence 4, Appl
41	62.5	16.9	2578	1	US-08-317-522A-4	Sequence 4, Appl
42	62	16.8	610	4	US-09-385-982-229	Sequence 229, App
43	62	16.8	1050	4	US-08-971-090-4	Sequence 4, Appl
44	62	16.8	6924	1	US-08-015-973-2	Sequence 2, Appl
45	62	16.8	6924	2	US-08-448-164-2	Sequence 2, Appl

RESULT 1

US-08-202-389-1

Sequence 1, Application US/08202389

Patent No. 5536636

GENERAL INFORMATION:

APPLICANT: Freeman Jr., Robert M.

APPLICANT: Plutsky, Jorge

APPLICANT: Neel, Benjamin G.

APPLICANT: Rosenbery, Robert D.

TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE

TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS

NUMBER OF SEQUENCES: 54

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Millia Drive

CITY: Lexington

STATE: MA

COUNTRY: USA

Zip: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/202,389

FILING DATE: 28-FEB-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/983,926

FILING DATE: 01-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/829,141

FILING DATE: 31-JAN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/721,112

FILING DATE: 26-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BIH92-05MA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1747 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1540
US-08-202-389-1

Alignment Scores:
Pred. No.: 0.00235 Length: 1747
Score: 88.00 Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.85% Indels: 4
Gaps: 2

US-09-935-703-7 (1-67) x US-08-202-389-1 (1-1747)

QY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18
Db 491 GAGGAGCTTTGAGAGTGTGCAAAAGCAAGAGTAGTAAGACTTGCCACCGCTGTG---GAA 547
QY 19 GlysGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
Db 548 GGGCAGCGGGCGGAGAGACAGACAGACAGACCGCTACAGAACATCTCCCTTTGACAC 607

RESULT 2
US-08-202-389-7
Sequence 7, Application US/08202389
Patent No. 5536636
GENERAL INFORMATION:
APPLICANT: Freeman Jr., Robert M.
APPLICANT: Plutsky, Jorge
APPLICANT: Neel, Benjamin G.
APPLICANT: Rosenberg, Robert D.
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE
TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,389
FILING DATE: 28-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/983,926
FILING DATE: 01-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/829,141
FILING DATE: 31-JAN-1992
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/721,112
FILING DATE: 26-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BIH92-05MA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2143 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 145..2037
US-08-202-389-7

Alignment Scores:
Pred. No.: 0.00624 Length: 2143
Score: 86.00 Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31% Indels: 4
Gaps: 2

US-09-935-703-7 (1-67) x US-08-202-389-7 (1-2143)

QY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18
Db 880 GAGGAGCTTTGAGAGTGTGCAAGAGCAGAGGTGAAGACTTGCCACCGCTGTG---GAA 936
QY 19 GlysGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
Db 937 GGGCAGCGGGCGGAGAGACAGAGGAGAGACCGCTACAGAACATCTCCCTTTGACAC 996

RESULT 3
US-08-202-389-5
Sequence 5, Application US/08202389
Patent No. 5536636
GENERAL INFORMATION:
APPLICANT: Freeman Jr., Robert M.
APPLICANT: Plutsky, Jorge
APPLICANT: Neel, Benjamin G.
APPLICANT: Rosenberg, Robert D.
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE
TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,389
FILING DATE: 28-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/983,926
FILING DATE: 01-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/829,141
FILING DATE: 31-JAN-1992
PRIOR APPLICATION DATA:

Db 907 GGTACCAAGAAAGAAATTCATGCTGCACAGACCAAGAAAGAAAGCGTGAAT 966
QY 50 ---Phetp 51
|||||
Db 967 GATTCTGG 975

RESULT 6

US-08-015-985-4
: Sequence 4, Application US/08015985
: Patent No. 5538886
: GENERAL INFORMATION:
: APPLICANT: Schlessinger, Joseph
: APPLICANT: Sap, Jan M.
: TITLE OF INVENTION: NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
: TITLE OF INVENTION: PHOSPHATASE-ALPHA
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 AVENUE OF THE AMERICAS
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: U.S.A.
: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/015,985
: FILING DATE: 10-FEB-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Coruzzi, Laura A.
: REGISTRATION NUMBER: 30,742
: REFERENCE/DOCKET NUMBER: 7683-020
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-9741/8864
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2872 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
US-08-015-985-4

Alignment Scores:

Pred. No.: 0.0875 Length: 2872
Score: 79.50 Matches: 22
Percent Similarity: 37.35% Conservative: 9
Best Local Similarity: 26.51% Mismatches: 19
Query Match: 21.54% Indels: 33
Db: 1 Gaps: 3

US-09-935-703-7 (1-67) x US-08-015-985-4 (1-2872)

QY 2 GINGLuphMetaLaleuLuleuLysasnLeuproGlyLupheasnSergLYasngln 21
Db 958 GAGGAAATTACCGCTCTGCTGCTTCTATCCAGGCCACCTGTGAGGCTGCCCTCAG 1017
QY 22 ProSerAsnArgGluLysAsnArgTyrArgAspIleLeuprobheglnHis 38
Db 1018 GAGGAAACAGGAAACCAACCGCTATGTAAACATCCCTGCTATGACCACTAGAGTG 1077
QY 39 -----His 39
Db 1078 CACCTGACACCTGTGAAGGGGTCCAGATTCGTATACATCAACGCTTCATCATTAAT 1137
QY 40 GLYTYR-----SergLYProAsnGlnLupgThr--- 49
||||| ||||| ||| |||

Db 1138 GGTACCAAGAAAGAAATTCATGCTGCACAGACCAAGAAAGAAAGCGTGAAT 1197
QY 50 ---Phetp 51
|||||
Db 1198 GACTTCTGG 1206

RESULT 7

US-08-821-278A-16
: Sequence 16, Application US/08821278A
: Patent No. 6238802
: GENERAL INFORMATION:
: APPLICANT: Cheng, Jili
: APPLICANT: Lasky, Laurence A.
: TITLE OF INVENTION: Protein Tyrosine Phosphatases
: FILE REFERENCE: P1010R1
: CURRENT APPLICATION NUMBER: US/08/821,278A
: CURRENT FILING DATE: 1997-03-20
: NUMBER OF SEQ ID NOS: 23
: SEQ ID NO 16
: LENGTH: 466
: TYPE: DNA
: ORGANISM: Homo Sapien
US-08-821-278A-16

Alignment Scores:

Pred. No.: 0.00913 Length: 466
Score: 78.50 Matches: 26
Percent Similarity: 33.96% Conservative: 10
Best Local Similarity: 24.53% Mismatches: 19
Query Match: 21.27% Indels: 51
Db: 4 Gaps: 4

US-09-935-703-7 (1-67) x US-08-821-278A-16 (1-466)

QY 12 LeuProGlyLupheasn----- 17
Db 28 CTCGCCGGAGGTTCAGCATCCAGGCGCTGCGCCCTGAGAGCTGACGGCGTG 87
QY 18 -----SergLYasnglnProSerAsnArgGluLysAsnArgTyrArgAspIle 33
Db 88 TGCTCACCGCTGCGCCGCGAGTCCGCGAGAGACGTGAGAGAACCCCTACAAAGAGCTG 147
QY 34 LeuProPhe-----GlnHisHisGlyTyrSer--- 42
Db 148 CTCGCTTATGATCAGACGCGAGTAATCCTCTCCCTGCTCCAGGAAGGACACAGCGAC 207
QY 42 ----- 42
Db 208 TACATTATGCACTTCATCCGGGCGTGATGGAAGCCTGCTACATTGCCACGCA 267
QY 43 GLYProAsnGluArgThr-----ThrPhetpHisGlySerAsnGlnGluAlaValSer 60
Db 268 GGACCCCTGCCACACACCCCTGCTAGACTTCTGAGACTGTGCGAGCTGTGGAGTTGGGTCAG 327
QY 61 LeuLeuLeuArgTyrCys 66
Db 328 GTGATCTGATGCGCTGT 345

RESULT 8
US-09-152-060-24
: Sequence 24, Application US/09152060
: Patent No. 6448230.
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: 28 Human Secreted Proteins
: FILE REFERENCE: P2003P1 US
: CURRENT APPLICATION NUMBER: US/09/152,060
: CURRENT FILING DATE: 1998-09-11
: EARLIER APPLICATION NUMBER: PCT/US98/04858
: EARLIER FILING DATE: 1998-03-12
: EARLIER APPLICATION NUMBER: 60/040,762
: EARLIER FILING DATE: 1997-03-14
: EARLIER APPLICATION NUMBER: 60/040,710

```

EARLIER FILING DATE: 1997-03-14
EARLIER APPLICATION NUMBER: 60/050,934
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/048,100
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/048,357
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/048,189
EARLIER FILING DATE: 1997-05-30
EARLIER APPLICATION NUMBER: 60/057,765
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: 60/048,970
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/068,368
EARLIER FILING DATE: 1997-12-19
NUMBER OF SEQ ID NOS: 118
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 24
LENGTH: 1659
TYPE: DNA
ORGANISM: Homo sapiens
US-09-152-060-24

Alignment Scores:
Pred. No.:      0.31      Length:      1659
Score:          73.50     Matches:      20
Percent Similarity: 50.94%   Conservative: 7
Best Local Similarity: 37.74% Mismatches:    15
Query Match:    19.92%     Indels:       11
                        Gaps:        2

US-935-703-7 (1-67) x US-09-152-060-24 (1-1659)

OY      9 LeuIysAsnLeuProGlycIuPheAsnSerGlyAsnGlnProSerAsnArgGluTyAsn 28
Db      84 CTTAGAGTCATCATCTCAAGAGACACACTCGCACAATTCCCTGC-----128
OY      29 ArgTyrGAspIleLeuProPheGlnHisGlyTyrSerIlyProAsn-----45
Db      129 -----AAGATTCACTTCCCTTGATTTGCCAAGGGGCATTGGGGCCTCCTTTGCAAAT 162

OY      46 --GUArGrThrPhetRphIsGlSerAsnGluGly 57
Db      183 GTAGAGGAGCATTTCTTTCAGCTATGAGAGCTAGAGACGA 221

RESULT 9
US-08-446-345-33
Sequence 33, Application US/08446345
Patent No. 5631009
GENERAL INFORMATION:
APPLICANT: Ullrich, Axel
APPLICANT: Moller, Niels P.H.
TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE
TITLE OF INVENTION: PHOSPHATASES PTP-D1
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: N.Y.
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,345
FILING DATE: 22-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

```

```

1  APPLICATION NUMBER: US 08/234,440
2  FILING DATE: 28-APR-1994
3  ATTORNEY/AGENT INFORMATION:
4  NAME: CORUZZI, Laura A.
5  REGISTRATION NUMBER: 30742
6  REFERENCE/DOCKET NUMBER: 7663-054
7  TELECOMMUNICATION INFORMATION:
8  TELEPHONE: (212) 790-9090
9  TELEFAX: (212) 869-8864
10  TELEX: 66141 PENNTE
11  INFORMATION FOR SEQ ID NO: 33:
12  SEQUENCE CHARACTERISTICS:
13  LENGTH: 878 base pairs
14  TYPE: nucleic acid
15  STRANDEDNESS: both
16  TOPOLOGY: unknown
17  MOLECULE TYPE: CDNA
18  US-08-446-345-33
19
20  Alignment Scores:
21  Pct. Ident. 0.148 Length: 878
22  Score: 73.00 Matches: 17
23  Percent Similarity: 59.57% Conservative: 11
24  Best Local Similarity: 36.17% Mismatches: 17
25  Query Match: 19.78% Indels: 2
26  DB: 2 Gaps: 0
27
28  US-09-935-703-7 (1-67) x US-08-446-345-33 (1-878)
29
30  QY 7 LeuGlulLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnInProSerAsnArgGlu 26
31  ||||: ||: ||: ||||| :||| ||||
32  DB 6 CTTAGAAACGGCTAGTGTATGCGGAGTGCTCAACAGCAGATCCCTGAAATGCAGAA 65
33
34  QY 27 LysAsnArgTyrArgAspIleLeuProPheGlnHis-HisGlyTyrSer-GlyProAsn 46
35  ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
36  DB 66 AGAAATCGATTCAGATGTCTTCTCTATGATGATGCAGAGTGGAGTGTGCCACT 125
37  ||||| ||||| |||||
38  QY 46 LuArgThrThrPheTyr 51
39  ||||| |||||
40  DB 126 AAAGAAAACACACTGG 142
41
42  RESULT 10
43  US-08-446-345-35
44  ; Sequence 35, Application US/08446345
45  ; Patent No. 5831009
46  ; GENERAL INFORMATION:
47  ; APPLICANT: Ollrich, Axel
48  ; APPLICANT: Mollet, Niels P.H.
49  ; APPLICANT: Mollet, Karin B.
50  ; TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE
51  ; BEST OF INVENTION: PHOSPHATASES PTP-D1
52  ; NUMBER OF SEQUENCES: 41
53  ; CORRESPONDENCE ADDRESS:
54  ; ADDRESSEE: Pennie & Edmonds
55  ; STREET: 1155 Avenue of the Americas
56  ; CITY: New York
57  ; STATE: N.Y.
58  ; COUNTRY: U.S.A.
59  ; ZIP: 10036-2711
60  ; COMPUTER READABLE FORM:
61  ; MEDIUM TYPE: Floppy disk
62  ; COMPUTER: IBM PC compatible
63  ; OPERATING SYSTEM: PC-DOS/MS-DOS
64  ; SOFTWARE: Patent Release #1.0, Version #1.25
65  ; CURRENT APPLICATION DATA:
66  ; APPLICATION NUMBER: US/08/446,345
67  ; FILING DATE: 22-MAY-1995
68  ; CLASSIFICATION: 435
69  ; PRIOR APPLICATION DATA:
70  ; APPLICATION NUMBER: US 08/234,440
71  ; FILING DATE: 28-APR-1994
72  ; ATTORNEY/AGENT INFORMATION:
73  ; NAME: Coruzzi, Laura A.

```

REGISTRATION NUMBER: 30742
REFERENCE/DOCKET NUMBER: 7683-054
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 4080 base pairs
STRANDEDNESS: both
TYPE: nucleic acid
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
US-08-446-345-35

Alignment Scores:
Pred. No.: 1.34 Length: 4080
Score: 73.00 Matches: 17
Percent Similarity: 59.57% Conservative: 11
Best Local Similarity: 36.17% Mismatches: 17
Query Match: 19.78% Indels: 2
DB: 2 Gaps: 0

US-09-935-703-7 (1-67) x US-08-446-345-35 (1-4080)

QY 7 LeuGlutLeuLysAsnLeuProGlyLupheAsnSerGlyAsnGlnProSerAsnArgIu 26
Db 3042 CTTAAGAAAGCGGCTGATGGGAGTGGCTCAACAGCAGCTCCCGAAATGCAGAA 3101

QY 27 LysAsnArgTyrArgAspIleLeuProPheGlnHis-HisGlyTyrSer-GlyProAsn 46
Db 3102 AGAAATCGATGCCAAGATGTTCTTCCTTATGATGATGCGAAGTGGAGTGTGCTCAACT 3161

QY 46 LuArgThrThrPheTrp 51
Db 3162 AAAGAAACACACTGG 3178

RESULT 11
US-09-358-683-1
Sequence 1, Application US/09358683
Patent No. 6200807
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
TITLE OF INVENTION: ANTISENSE MODULATION OF SHP-2 EXPRESSION
FILE REFERENCE: RTS-0082
CURRENT APPLICATION NUMBER: US/09/358, 683
CURRENT FILING DATE: 1999-07-21
NUMBER OF SEQ ID NOS: 47
SEQ ID NO 1
LENGTH: 2121
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (154)..(1935)
US-09-358-683-1

Alignment Scores:
Pred. No.: 0.737 Length: 2121
Score: 72.00 Matches: 16
Percent Similarity: 54.76% Conservative: 7
Best Local Similarity: 38.10% Mismatches: 13
Query Match: 19.51% Indels: 6
DB: 4 Gaps: 2

US-09-935-703-7 (1-67) x US-09-358-683-1 (1-2121)

QY 19 GlyAsnGlnProSerAsnArgGlyLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
Db 955 GGTCAAGGCAAGCAAAACAAACAAATAGATTAATAAAATCCTGCTTGATCAT 1014

QY 39 -----HisGlyTyrSerGlyProAsnGluArgThrThrPheTrpHisGly 53

Db 1015 ACCAGGTTGCTTACACGAT---GGTATCCCAATGAGCCTGTTTCAGATTACTCAAT 1071
QY 54 SerAsn 55
Db 1072 GCAAT 1077

RESULT 12
US-08-202-389-11

Sequence 11, Application US/08202389
Patent No. 553636
GENERAL INFORMATION:
APPLICANT: Freeman Jr., Robert M.
APPLICANT: Plutsky, Jorge
APPLICANT: Neel, Benjamin G.
APPLICANT: Rosenberg, Robert D.
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,389
FILING DATE: 28-FEB-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/983,926
FILING DATE: 01-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/829,141
FILING DATE: 31-JAN-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/721,112
FILING DATE: 26-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BIR92-05MA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2276 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 114..1893
US-08-202-389-11

Alignment Scores:
Pred. No.: 0.815 Length: 2276
Score: 72.00 Matches: 16
Percent Similarity: 54.76% Conservative: 7
Best Local Similarity: 38.10% Mismatches: 13
Query Match: 19.51% Indels: 6
DB: 1 Gaps: 2

US-09-935-703-7 (1-67) x US-08-202-389-11 (1-2276)

```

QY      19 GJLYANGngInProSerAsnArgGluLysAsnArgTYRATgAspIleuenuProheginHs 38
      ||| ::|  ||| ::|  ||| ::|  ||| ::|  ||| ::|  ||| ::|  ||| ::|  |||
DB      915 GGTCAAGACGACAGAAAACAAAAACAAAATAGATATATAAAACATCTGCCCTTTGATCAT 974
QY      39 -----HisGlyTyISergIyProaIngIuArgThrThpHeTrpHisgly 53
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB      975 ACCAGGGTTGCTCTACAGAT---GGTGAATCCCAATGAGCCGTTCATACATTAATCAAT 1032
      1032 GCAAAT 1037
      US-08-594-031-74
      Sequence 74, Application US/08594031
      Patent No. 5783182
      GENERAL INFORMATION:
      APPLICANT: THOMPSON, Timothy C.
      TITLE OF INVENTION: METHOD FOR IDENTIFYING METASTATIC SEQUENCES
      NUMBER OF SEQUENCES: 175
      CORRESPONDENCE ADDRESS:
      ADDRESSEE: BAKER & BOTTS, L.L.P.
      STREET: 1299 Pennsylvania Avenue, N.W.
      CITY: Washington
      STATE: DC
      COUNTRY: USA
      ZIP: 20004-2400
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Diskette
      COMPUTER: IBM Compatible
      OPERATING SYSTEM: DOS
      SOFTWARE: FASTSEQ Version 1.5
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/594,031
      FILING DATE: 30-JAN-1996
      CLASSIFICATION: 435
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 60/006,838
      FILING DATE: 16-NOV-1995
      ATTORNEY/AGENT INFORMATION:
      NAME: Remenick, James
      REGISTRATION NUMBER: 36,902
      REFERENCE/DOCKET NUMBER: 0A146-0110
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 202-639-7700
      TELEFAX: 202-639-7890
      TELEX:
      INFORMATION FOR SEQ ID NO: 74:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 247 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: cDNA
      HYPOTHETICAL: NO
      ANTI-SENSE: NO
      FRAGMENT TYPE:
      ORIGINAL SOURCE:
      US-08-594-031-74
      Alignment Scores:
      Pred. No.: 0.0478 Length: 247
      Score: 71.00 Matches: 12
      Percent Similarity: 76.19% Conservative: 4
      Best Local Similarity: 57.14% Mismatches: 5
      Query Match: 19.24% Indels: 0
      DB: 1 Gaps: 0
      US-09-935-703-7 (1-67) x US-08-594-031-74 (1-247)
QY      18 SerGIyAngInProSerAsnArgGluLysAsnArgTYRATgAspIleuenuProhegin 37
      ||| ::|  ||| ::|  ||| ::|  ||| ::|  ||| ::|  ||| ::|  |||
DB      72 ACTGAGAGAAAGAGAGAAATGTTTAAAGAAACAGATATTAAGACATCTGCAATTGAT 131

```

```
QY      38 HLS 38  
       |||  
Db      132 CAC 134
```

RESULT 14
US-09-280-116-218/c
; Sequence 218, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 218
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2
US-09-280-116-218

Alignment Scores:

Pred. No.:	0.247	Length:	612
Score:	70.00	Matches:	20
Percent Similarity:	45.45%	Conservative:	10
Best Local Similarity:	30.30%	Mismatches:	16
Query Match:	18.97%	Indels:	20
DB:	4	Gaps:	4

US-09-935-703-7 (1-67) x US-09-280-116-218 (1-612)

QY 4 PheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGlnProSer 23
 |||||:||||| ||| :||| ||| :||| |||||
Db 397 TTCTTGCGCAGAGAGAGCAAGAAGCTCCGCC-----AGTTCGCATCA 356

QY 24 AsnaTgGluLysAsnArg-----TyraTgaSpIleLeuProPheGlnHisGlyTyr 41
 ||||| ||| ||| :||| ||| |||
Db 355 TCAGCGGAATCCTTCGGCGGGGTTCACAGAGCCCTCCCAACTTATCCAC----- 305

QY 42 SerGlyProAsnGlnArgThrPhetRphHisGlySerAsnGlnGlyAlaValSerIeu 61
 :|||:|||||:|||||:|||||:|||||:|||||
Db 304 -----TATTGGCATGGTGCATCACTGGAGTGTGCAGACTTG 269

QY 62 ---LeuLeuArgTyrCys 66
 ||||| ||||| |||||
Db 268 GGCCCTGCTGTCTCTACTGC 251

RESULT 15
US-08-018-129-4
; Sequence 4, Application US/08018129
; Patent No. 5589375
; GENERAL INFORMATION:
; APPLICANT: Ullrich, Axel
; TITLE OF INVENTION: PTP ID: A NOVEL PROTEIN TYROSINE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patentin release #1.0, version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/018.129
 FILING DATE: 19930216
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Mistrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7683-017
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2790 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 130..1911
 US-08-018-129-4

Alignment Scores:
 Pred. No.: 2.16
 Score: 70.00
 Percent Similarity: 53.66%
 Best Local Similarity: 36.59%
 Query Match: 18.97%
 DB: 1
 Length: 2790
 Matches: 15
 Conservative: 7
 Mismatches: 15
 Indels: 4
 Gaps: 1

US-09-935-703-7 (1-67) x US-08-018-129-4 (1-2790)

QY 19 GlyAsnGlnProSerAsnArgGluLyAsnArgTyrArgAspIleLeuProPheGlnHis 38
 DB 931 GGTCAAGGCGCAGCAAAACAAATAAGATATAAAACATCCCTGCGCTTGATCAT 990
 QY 39 Hls-----GlyTyrSerGlyProAsnGluArgThrThrPheTrpHisGlySer 54
 DB 991 ACCAGGTTGTCTCTCAGCATCTGTGATCCCAATGAGCGCTGTTTCAGATTCATCATGCA 1050
 QY 55 Asn 55
 DB 1051 AAT 1053

Search completed: June 11, 2003, 14:05:54
 Job time : 71 secs

Thu Jun 12 08:30:30 2003

us-09-935-703-7.p2n.rnpb

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - nucleic search, using frame_plus.p2n model

Run on: June 11, 2003, 13:40:28 ; Search time 117 Seconds

(Without alignments)
801.444 Million cell updates/sec

Title: US-09-935-703-7

Perfect score: 369
Sequence: 1 MOEFMALELNLPDEFNSGN.....TTFHNSGVALIRYCA 67

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 870385 seqs, 699768693 residues

Total number of hits satisfying chosen parameters: 1740770

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame.p2n.model -DEV-xip
-O/cgcn2.1/USPTO.spool/US0935703/runtat_06062003-095213-6265/app_query.fasta.1.263
-DB-published.Applications.NA -OPT-fastap -SUFFIX-p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-blosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE-LOCAL -OUTFMT-pio -NORM-ext -HAPPSTZ=500 -MINLEN=0
-MAXLEN=200000000 -USER-US09935703.ecgn.1.1.80 -runtat_06062003-095213-6265
-NCP=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100
-LONGLOG -DEV-TIMEOUT=120 -WARN-TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FAPOP=6 -Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published.Applications.NA.*

1: /cgcn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgcn2.6/ptodata/1/pubpna/PTCT_NEW_PUB.seq.*
3: /cgcn2.6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgcn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
5: /cgcn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
6: /cgcn2.6/ptodata/1/pubpna/PTCTUS_PUBCOMB.seq.*
7: /cgcn2.6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgcn2.6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgcn2.6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgcn2.6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgcn2.6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgcn2.6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgcn2.6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgcn2.6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88.5	24.0	2389	9	US-10-098-841-192
2	88	23.8	402	10	US-09-960-352-3199
3	86	23.3	326	9	US-09-796-692-3635
4	86	23.3	326	9	US-09-796-692-3430
					Sequence 192, App
					Sequence 3199, Ap
					Sequence 3635, Ap
					Sequence 5490, Ap

5	86	23.3	326	9	US-10-040-862-3635	Seq.
6	86	23.3	326	9	US-10-040-862-5490	Sequence-
7	86	23.3	333	9	US-09-796-692-3721	Sequence 3721,
8	86	23.3	333	9	US-10-040-862-3721	Sequence 3721,
9	86	23.3	2076	10	US-09-822-849A-5	Sequence 5, Appl1
10	86	23.3	2145	10	US-09-920-021A-2	Sequence 2, Appl1
11	86	23.3	2176	10	US-09-917-800A-1478	Sequence 1478, Ap
12	86	23.3	2532	9	US-10-072-036-116	Sequence 116, Ap
13	86	23.3	2562	9	US-10-072-036-118	Sequence 118, App
14	86	23.3	3220	12	US-10-002-600-96	Sequence 36, Appl
15	81.5	22.1	535	9	US-09-918-985-31823	Sequence 31823, A
16	81.5	22.1	14738	9	US-09-764-891-5637	Sequence 5637, Ap
17	81.5	22.1	14738	9	US-09-764-891-5637	Sequence 5639, Ap
18	79.5	21.5	564	9	US-09-918-995-30413	Sequence 30413, A
19	78.5	21.3	2810	9	US-10-087-993-35	Sequence 35, Appl
20	73.5	19.9	488	10	US-09-864-761-11796	Sequence 11796, A
21	73.5	19.9	1659	9	US-09-852-797-24	Sequence 24, Appl
22	73.5	19.9	1659	10	US-09-853-161-24	Sequence 24, Appl
23	73.5	19.9	1659	10	US-09-852-659A-24	Sequence 24, Appl
24	73.5	19.9	4080	10	US-09-962-436-342	Sequence 342, App
25	72	19.5	1794	9	US-10-038-010-21	Sequence 21, Appl
26	72	19.5	2276	10	US-09-920-021A-4	Sequence 4, Appl
27	72	19.5	2300	12	US-10-044-090-622	Sequence 622, App
28	70	19.0	547	10	US-09-925-300-300	Sequence 300, App
29	70	19.0	1494	10	US-09-917-800A-1503	Sequence 1503, App
30	70	19.0	1551	9	US-10-180-921-1	Sequence 1, Appl1
31	68.5	18.6	295	9	US-09-736-457-1375	Sequence 1375, Ap
32	68.5	18.6	295	9	US-09-902-941-1375	Sequence 1375, Ap
33	68.5	18.6	295	9	US-09-849-626-1375	Sequence 1375, Ap
34	68.5	18.6	295	9	US-10-017-754-1375	Sequence 1375, Ap
35	68.5	18.6	3630	9	US-09-927-827-31	Sequence 31, Appl
36	68	18.4	2346	10	US-09-854-883-241	Sequence 241, App
37	68	18.4	4127	10	US-09-854-883-3	Sequence 3, Appl1
38	68	18.4	4127	10	US-09-854-883-10	Sequence 10, Appl1
39	68	18.4	4127	10	US-09-917-800A-1555	Sequence 1555, Ap
40	67.5	18.3	446	10	US-09-867-701-7967	Sequence 7967, Ap
41	67.5	18.3	2226	9	US-10-087-993-31	Sequence 31, Appl
42	67.5	18.3	2226	9	US-10-243-687-1	Sequence 1, Appl1
43	64.5	17.5	3580	10	US-09-822-295-1	Sequence 1, Appl1
44	64.5	17.5	3580	12	US-10-005-467-1	Sequence 1, Appl1
45	64.5	17.5	143306	10	US-09-729-920-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-10-098-841-192
; Sequence 192, Application US/10098841
; Publication No. US20020197679A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Xu, Chongjun
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungling
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhao, Qing A.
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: Weinman, Tom
APPLICANT: Zhang, Jie
APPLICANT: Qian, Xiaohong B.
TITLE OF INVENTION: Drmanac, Radote T.
TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784C1P2
CURRENT APPLICATION NUMBER: US/10/098,841
CURRENT FILING DATE: 2002-03-13
PRIOR APPLICATION NUMBER: 09/598,042
PRIOR FILING DATE: 2000-06-20

DB 189 GGGCAGCGGCCAGAGAACAGGCGACACCGCTACAGAACATTCTCCCTTTGACGAC 248

RESULT 4

US-09-796-692-5490

Sequence 5490, Application US/09796692

Publication No. US20020198362A1

GENERAL INFORMATION:

APPLICANT: Gaiger, Alexander

APPLICANT: Algate, Paul A.

APPLICANT: Mannion, Jane

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

FILE REFERENCE: 2077.001200

CURRENT APPLICATION NUMBER: US/09/796,692

PRIOR FILING DATE: 2001-03-01

PRIOR APPLICATION NUMBER: 60/186,126

PRIOR FILING DATE: 2000-03-01

PRIOR APPLICATION NUMBER: 60/190,479

PRIOR FILING DATE: 2000-03-17

PRIOR APPLICATION NUMBER: 60/200,545

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: 60/200,303

PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: 60/200,779

PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: 60/200,999

PRIOR FILING DATE: 2000-05-01

PRIOR APPLICATION NUMBER: 60/202,084

PRIOR FILING DATE: 2000-05-04

PRIOR APPLICATION NUMBER: 60/206,201

PRIOR FILING DATE: 2000-05-22

PRIOR APPLICATION NUMBER: 60/218,950

PRIOR FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: 60/222,903

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: 60/223,416

PRIOR FILING DATE: 2000-08-04

PRIOR APPLICATION NUMBER: 60/223,378

PRIOR FILING DATE: 2000-08-07

NUMBER OF SEQ ID NOS: 9597

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 5490

LENGTH: 326

TYPE: DNA

ORGANISM: Homo sapiens

US-09-796-692-5490

Alignment Scores:

Pred. No.:	0.000424	Length:	326
Score:	86.00	Matches:	19
Percent Similarity:	65.00%	Conservative:	7
Best Local Similarity:	47.50%	Mismatches:	10
Query Match:	23.31%	Indels:	4
DB:	9	Gaps:	2

US-09-935-703-7 (1-67) x US-09-796-692-5490 (1-326)

OY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18

DB 132 GAGGAGTTTGAGAGTTTGAGAGGAGGAGTGTAACACTTGACACGCTGTG---GAA 188

OY 19 GlyAsnGlnProSerAsnArgLysAsnArgTyrArgAspIleLeuProGlnHis 38

DB 189 GGGCAGCGGCCAGAGAACAGGCGACACCGCTACAGAACATTCTCCCTTTGACGAC 248

RESULT 5

US-10-040-862-3635

Sequence 3635, Application US/10040862

Publication No. US20030078396A1

GENERAL INFORMATION:

APPLICANT: Gaiger, Alexander

APPLICANT: Algate, Paul A.

APPLICANT: Mannion, Jane

APPLICANT: Retter, Marc

APPLICANT: Corixa Corporation

TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther

FILE REFERENCE: 014058-013520US

CURRENT APPLICATION NUMBER: US/10/040,862

CURRENT FILING DATE: 2001-11-06

PRIOR FILING DATE: 2000-03-01

PRIOR APPLICATION NUMBER: US 60/186,126

PRIOR FILING DATE: 2000-03-17

PRIOR APPLICATION NUMBER: US 60/190,479

PRIOR FILING DATE: 2000-03-17

PRIOR APPLICATION NUMBER: US 60/200,545

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: US 60/200,303

PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: US 60/200,779

PRIOR FILING DATE: 2000-04-28

PRIOR APPLICATION NUMBER: US 60/200,999

PRIOR FILING DATE: 2000-05-01

PRIOR APPLICATION NUMBER: US 60/202,084

PRIOR FILING DATE: 2000-05-04

PRIOR APPLICATION NUMBER: US 60/206,201

PRIOR FILING DATE: 2000-05-22

PRIOR APPLICATION NUMBER: US 60/218,950

PRIOR FILING DATE: 2000-07-14

PRIOR APPLICATION NUMBER: US 60/222,903

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: US 60/223,416

PRIOR FILING DATE: 2000-08-04

PRIOR APPLICATION NUMBER: US 60/223,378

PRIOR FILING DATE: 2000-08-07

PRIOR APPLICATION NUMBER: US 09/796,692

PRIOR FILING DATE: 2001-03-01

NUMBER OF SEQ ID NOS: 10467

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 3635

LENGTH: 326

TYPE: DNA

ORGANISM: Homo sapiens

US-10-040-862-3635

Alignment Scores:

Pred. No.:	0.000424	Length:	326
Score:	86.00	Matches:	19
Percent Similarity:	65.00%	Conservative:	7
Best Local Similarity:	47.50%	Mismatches:	10
Query Match:	23.31%	Indels:	4
DB:	9	Gaps:	2

US-09-935-703-7 (1-67) x US-10-040-862-3635 (1-326)

OY 2 GlnGluPheMetAlaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18

DB 132 GAGGAGTTTGAGAGTTTGAGAGGAGGAGTGTAACACTTGACACGCTGTG---GAA 188

OY 19 GlyAsnGlnProSerAsnArgLysAsnArgTyrArgAspIleLeuProGlnHis 38

DB 189 GGGCAGCGGCCAGAGAACAGGCGACACCGCTACAGAACATTCTCCCTTTGACGAC 248

RESULT 6

US-10-040-862-5490

Sequence 5490, Application US/10040862

Publication No. US20030078396A1

GENERAL INFORMATION:

APPLICANT: Gaiger, Alexander

APPLICANT: Algate, Paul A.

APPLICANT: Mannion, Jane

APPLICANT: Retter, Marc

TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther

FILE REFERENCE: 014058-013520US

CURRENT APPLICATION NUMBER: US/10/040,862

```

; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5490
; LENGTH: 326
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-5490

Alignment Scores:
Pred. No.: 0.000424      Length: 326
Score: 86.00             Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31%      Indels: 4
DB: 9                    Gaps: 2

US-09-935-703-7 (1-67) x US-10-040-862-5490 (1-326)

QY 2 GlnGluphemeMetaLaleu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18
Db 132 GAGGAGTTGAGAGTTTGCAGAGCAGAGGAGTGAGACTTGCACCGCTGTG--GAA 188
QY 19 GlysAngInProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
Db 189 GGGCAGCGCGCCAGACAGACAGAGGCGCAAGAACCGCTACAGAACATCTCCCTTTGACAC 248

RESULT 7
; Sequence 3721, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR APPLICATION NUMBER: 2001-03-01
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,303
```

```

; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3721
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)
; OTHER INFORMATION: n=A,T,C or G
US-09-796-692-3721

Alignment Scores:
Pred. No.: 0.000436      Length: 333
Score: 86.00             Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31%      Indels: 4
DB: 9                    Gaps: 2

US-09-935-703-7 (1-67) x US-09-796-692-3721 (1-333)

QY 2 GlnGluphemeMetaLaleu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18
Db 139 GAGGAGTTGAGAGTTTGCAGAGCAGAGGAGTGAGACTTGCACCGCTGTG--GAA 195
QY 19 GlysAngInProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
Db 196 GGGCAGCGCGCCAGACAGACAGAGGCGCAAGAACCGCTACAGAACATCTCCCTTTGACAC 255

RESULT 8
; Sequence 3721, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther
; FILE REFERENCE: 014058-01352005
; CURRENT APPLICATION NUMBER: US/10/040,862
; PRIOR APPLICATION NUMBER: 2001-11-06
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR APPLICATION NUMBER: US 60/200,999
```

```

; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3721
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)
; OTHER INFORMATION: n-A,T,C or G
US-10-040-862-3721

Alignment Scores:
Pred. No.: 0.000436 Length: 333
Score: 86.00 Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31% Indels: 4
DB: 9 Gaps: 2

US-09-935-703-7 (1-67) x US-10-040-862-3721 (1-333)

QY 2 GlnGluPheMetAlaLeu-----GluLeuYsaNleuProGlyGluPheAsnSer 18
Db 139 GAGGAGTTGAGAGTTTCAGAGACGAGAGGTGAAGAACTTGACCAAGCGCTGTG---GAA 195

QY 19 GlyAsnGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
Db 196 GGGCAGCGCCGACAGAACAGGCAAGGCAAGCCGCTACAGAACATTTCTCCCTTTGACAC 255

RESULT 9
US-09-822-849A-5
; Sequence 5, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakari
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 2076
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-5
```

```

Alignment Scores:
Pred. No.: 0.00494 Length: 2076
Score: 86.00 Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31% Indels: 4
DB: 10 Gaps: 2

US-09-935-703-7 (1-67) x US-09-822-849A-5 (1-2076)

QY 2 GlnGluPheMetAlaLeu-----GluLeuYsaNleuProGlyGluPheAsnSer 18
Db 808 GAGGAGTTGAGAGTTTCAGAGACGAGAGGTGAAGAACTTGACCAAGCGCTGTG---GAA 864

QY 19 GlyAsnGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
Db 865 GGGCAGCGCCGACAGAACAGGCAAGGCAAGCCGCTACAGAACATTTCTCCCTTTGACAC 924

RESULT 10
US-09-920-021A-2
; Sequence 2, Application US/09920021A
; Patent No. US20020110800A1
; GENERAL INFORMATION:
; APPLICANT: MARSH, H. NICK
; APPLICANT: KAPLAN, DAVID
; TITLE OF INVENTION: USE OF SHP-1 AND SHP-2 TO DETECT
; FILE REFERENCE: 08338/016001
; CURRENT APPLICATION NUMBER: US/09/920,021A
; PRIOR FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: US/08/918,157
; PRIOR FILING DATE: 1997-08-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2145
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-920-021A-2

Alignment Scores:
Pred. No.: 0.00516 Length: 2145
Score: 86.00 Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31% Indels: 4
DB: 10 Gaps: 2

US-09-935-703-7 (1-67) x US-09-920-021A-2 (1-2145)

QY 2 GlnGluPheMetAlaLeu-----GluLeuYsaNleuProGlyGluPheAsnSer 18
Db 880 GAGGAGTTGAGAGTTTCAGAGACGAGAGGTGAAGAACTTGACCAAGCGCTGTG---GAA 936

QY 19 GlyAsnGlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPheGlnHis 38
Db 937 GGGCAGCGCCGACAGAACAGGCAAGGCAAGCCGCTACAGAACATTTCTCCCTTTGACAC 996

RESULT 11
US-09-917-800A-1478
; Sequence 1478, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
```

PRIOR APPLICATION NUMBER: US 60/222,040
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: US 60/222,880
PRIOR FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: US 60/290,029
PRIOR FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: US 60/290,645
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: US 60/292,336
PRIOR FILING DATE: 2001-05-22
PRIOR APPLICATION NUMBER: US 60/295,798
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: US 60/297,457
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,884
PRIOR FILING DATE: 2001-06-19
PRIOR APPLICATION NUMBER: US 60/303,459
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 1740
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1478
LENGTH: 2176
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020119462A1 U77038
US-09-917-800A-1478

Alignment Scores:
Pred. No.: 0.00526 Length: 2176
Score: 86.00 Matches: 19
Percent Similarity: 62.50% Conservative: 6
Best Local Similarity: 47.50% Mismatches: 11
Query Match: 23.31% Indels: 4
Gaps: 2

US-09-935-703-7 (1-67) x US-09-917-800A-1478 (1-2176)

QY 2 GlnGluPheMetAlaLeu-----GluLeuYsAnLeuProGlyGluPheAnsSer 18
Db 921 GAGGAGTTTGAGACTTTCGCAAAAGCAAGAGCGCAAGACTTGCACCGCTCTG--GAA 977
QY 19 GlysngInProSeAsnArGluLysAsnArGlyTyrArGAsPileuProPheGlnHis 38
Db 978 GGGCAGCGGCCGAGAGAACAGAGCAAGACCGCTACAGAACATCTCTCCCTTGACCAC 1037

RESULT 12
US-10-072-036-116
Sequence 116, Application US/10072036
Publication No. US20030082564A1
GENERAL INFORMATION:
APPLICANT: Ole THASTRUP
APPLICANT: Sara BJORN
APPLICANT: Soren TULLIN
APPLICANT: Kasper ALMHOLT
APPLICANT: Kurt SCUDDER
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
FILE REFERENCE: 3759-0120P
CURRENT APPLICATION NUMBER: US/10/072,036
PRIOR FILING DATE: 2002-09-13
PRIOR APPLICATION NUMBER: 09/417,197
NUMBER OF SEQ ID NOS: 143
SOFTWARE: PatentIn version 3.0
SEQ ID NO 116
LENGTH: 2532
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: EGFP-PTP fusion
NAME/KEY: CDS

LOCATOR: (1)..(2529)
US-10-072-036-116

Alignment Scores:
Pred. No.: 0.00643 Length: 2532
Score: 86.00 Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31% Indels: 4
Gaps: 2

US-09-935-703-7 (1-67) x US-10-072-036-116 (1-2532)

QY 2 GlnGluPheMetAlaLeu-----GluLeuYsAnLeuProGlyGluPheAnsSer 18
Db 1480 GAGGAGTTTGAGAGTTTCAGAGCAGAGAGTGAGAACTTGCACCGCTCTG--GAA 1536
QY 19 GlysngInProSeAsnArGluLysAsnArGlyTyrArGAsPileuProPheGlnHis 38
Db 1537 GGGCAGCGGCCGAGAGAACAGAGCGCTACAGAACATCTCTCCCTTGACCAC 1596

RESULT 13
US-10-072-036-118
Sequence 118, Application US/10072036
Publication No. US20030082564A1
GENERAL INFORMATION:
APPLICANT: Ole THASTRUP
APPLICANT: Sara BJORN
APPLICANT: Soren TULLIN
APPLICANT: Kasper ALMHOLT
APPLICANT: Kurt SCUDDER
TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To A
FILE REFERENCE: 3759-0120P
CURRENT APPLICATION NUMBER: US/10/072,036
PRIOR FILING DATE: 2002-09-13
PRIOR APPLICATION NUMBER: 09/417,197
NUMBER OF SEQ ID NOS: 143
SOFTWARE: PatentIn version 3.0
SEQ ID NO 118
LENGTH: 2562
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PTP-EGFP fusion
NAME/KEY: CDS
LOCATOR: (1)..(2559)
US-10-072-036-118

Alignment Scores:
Pred. No.: 0.00653 Length: 2562
Score: 86.00 Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31% Indels: 4
Gaps: 2

US-09-935-703-7 (1-67) x US-10-072-036-118 (1-2562)

QY 2 GlnGluPheMetAlaLeu-----GluLeuYsAnLeuProGlyGluPheAnsSer 18
Db 742 GAGGAGTTTGAGACTTTCGCAAGCAGAGGTGAAGACTTGCACCGCTCTG--GAA 798
QY 19 GlysngInProSeAsnArGluLysAsnArGlyTyrArGAsPileuProPheGlnHis 38
Db 799 GGGCAGCGGCCGAGAGAACAGAGCGCTACAGAACATCTCTCCCTTGACCAC 858

RESULT 14
US-10-002-600-96
Sequence 96, Application US/10002600
Patent No. US20020137077A1

```

; GENERAL INFORMATION:
; APPLICANT: Hopkins, Christopher M.
; APPLICANT: Peterson, David P.
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: GENES REGULATED IN ACTIVATED T CELLS
; FILE REFERENCE: PA-0042 US
; CURRENT APPLICATION NUMBER: US/10/002,600
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 60/243,521
; PRIOR FILING DATE: 2000-10-25
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 96
; LENGTH: 3220
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Template ID: 293296.19
US-10-002-600-96

Alignment Scores:
Pred. No.: 0.00884 Length: 3220
Score: 86.00 Matches: 19
Percent Similarity: 65.00% Conservative: 7
Best Local Similarity: 47.50% Mismatches: 10
Query Match: 23.31% Indels: 4
DB: 12 Gaps: 2

US-09-935-703-7 (1-67) x US-10-002-600-96 (1-3220)
Oy 2 GlngluphemeTalaLeu-----GluLeuLysAsnLeuProGlyGluPheAsnSer 18
   ||:||||| ||:|||||
Db 924 GAGGAGTTGAGAGTTGCGAAGCAGGAGGTGAAGACTTGACACCGCTCG---GAA 980
   ||:||||| ||:|||||
Oy 19 GlyaAngInProSerAsnArgGluLysAsnArgTyrArgAspIleuProPheGlnHis 38
   ||:||||| ||:|||||
Db 981 GGGCAGCGCGCCAGAGAACAGGCAAGAACCCCTACAGAACATCTCCCTTTGACCCAC 1040
   ||:||||| ||:|||||

RESULT 15
US-09-918-995-31823
; Sequence 31823, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31823
; LENGTH: 535
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(535)
; OTHER INFORMATION: n - A,T,C or G
US-09-918-995-31823

Alignment Scores:
Pred. No.: 0.00392 Length: 535
Score: 81.50 Matches: 16
Percent Similarity: 67.74% Conservative: 5
Best Local Similarity: 51.61% Mismatches: 9
Query Match: 22.09% Indels: 1
DB: 9 Gaps: 1

```

```

US-09-935-703-7 (1-67) x US-09-918-995-31823 (1-535)
Oy 8 GluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsnGlnProSerAsnArgGluLys 2'
   ||:||||| ||:|||||
Db 51 GAGGTGAAGAACTTGACACCGCTCG---GAAGGCGAGCGGCCAGAGAACAGGCGCAG 107
   ||:||||| ||:|||||
Oy 28 AsnArgTyrArgAspIleuProPheGlnHis 38
   ||:||||| ||:|||||
Db 108 AACCGCTACAAAGAACATCTCCCTTTGACCCAC 140
   ||:||||| ||:|||||

Search completed: June 11, 2003, 14:56:14
Job time : 120 secs

```

Gencore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: June 11, 2003, 11:46:42 ; Search time 1170 Seconds

(without alignments)
1666.571 Million cell updates/sec

Title: US-09-935-703-7

Sequence: 1 MGFMALEKMLPGEFNSGN.....TTFMGSGNGAVSLIRYCA 67

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09935703/runat_06062003_095211_6154/app-query.fasta.1.263
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blomsun62 -TRANS=human40.ccl -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09935703 @cgn.1.1.3745 @runat_06062003_095211_6154 -NCPU=6 -ICPU=3
-NO_MMAR -LARGOQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vt:*
15: gb_vt:*
16: em_da:*
17: em_fun:*
18: em_hum:*
19: em_in:*
20: em_mu:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	369	100.0	2499	9	BC036539
2	239.5	64.9	1914	6	AX354535
3	231.5	62.7	2786	9	HSMB00374
C 4	159	43.1	103737	9	AL603966
C 5	159	43.1	171930	9	AL450334
C 6	159	43.1	197389	9	AL358791
C 7	138	37.4	65229	9	AL672108
8	123	33.3	3090	10	D64141
9	88.5	24.0	892	10	RATPTEB
10	88.5	24.0	2155	10	RATPTECA
11	88.5	24.0	2160	9	HSPTPE
12	88.5	24.0	2209	9	HSJA315969
13	88.5	24.0	2210	10	MMU40280
14	88.5	24.0	2292	10	MMU62387
15	88.5	24.0	2827	10	MUSPTPE
16	88.5	24.0	3031	9	HSJA30580
17	88.5	24.0	4979	10	MMU36758
18	88.5	24.0	5425	6	AX305420
19	88.5	24.0	5425	10	MMU35368
20	88	23.8	1747	6	I23556
21	88	23.8	2122	10	MUSPTPEB
22	88	23.8	2163	10	MUSPTPEA
23	88	23.8	2168	10	BC012660
24	87.5	23.7	173224	2	AC130106
25	86	23.3	2090	9	HGMHCP
26	86	23.3	2121	9	AF178946
27	86	23.3	2143	6	I23559
28	86	23.3	2145	6	I23558
29	86	23.3	2145	6	I23558
30	86	23.3	2145	6	I23558
31	86	23.3	2176	6	AX401802
32	86	23.3	2190	10	RNU77038
33	86	23.3	2229	9	BC007667
34	86	23.3	2277	9	BC002523
35	86	23.3	2277	9	HSPTPEIC
36	86	23.3	2532	6	A84501
37	86	23.3	2532	6	A84503
38	86	23.3	2562	6	AX427400
39	85	23.0	2741	3	AX427402
40	83.5	22.6	1840	10	AF015682
41	83	22.5	2416	5	MMSHP1503
42	81.5	22.1	1099	9	XIUI1287
43	81.5	22.1	157115	9	HSPTPEICG04
44	81.5	22.1	222930	9	AC006512
45	81	22.0	25643	3	HSU47924

RESULT 1

ALIGNMENTS

BC036539 Homo sapi
AX354535 Sequence
AL050040 Homo sapi
AL603966 Human DNA
AL450334 Human DNA
AL358791 Human DNA
AL672108 Human DNA
D64141 Mus musculus
D78613 Rat mRNA fo
D78610 Rat mRNA fo
X54134 Human HPTP
AJ315969 Homo sapi
U40280 Mus musculus
U62387 Mus musculus
D83488 Mouse mRNA
AJ30580 Homo sapi
U36758 Mus musculus
AX305420 Sequence
U35368 Mus musculus
I23556 Sequence 1
M80389 Mouse prote
M88902 Mus musculu
BC012660 Mus muscu
AC130106 Rattus no
M74903 Human hemat
AF178946 Homo sapi
I23559 Sequence 7
I23558 Sequence 5
M77273 Human tyros
AX401802 Sequence
U77038 Rattus norv
BC007667 Homo sapi
BC002523 Homo sapi
X62055 H. sapiens p
A84501 Sequence 11
AX427400 Sequence 11
A84503 Sequence 11
AF015682 Sequence
AF015682 Caenorhab
U65953 Mus musculu
U15287 Xenopus lae
U15534 Human prote
AC006512 Homo sapi
U47924 Human chrom
U53332 Caenorhabd1

BC036539
 LOCUS BC036539 2499 bp mRNA linear PRI 20-AUG-2002
 DEFINITION Homo sapiens, similar to protein-tyrosine-phosphatase homolog DKFZP566K0524.1 - human (fragment), clone MGC:33863 IMAGE:5262458.
 mRNA, complete cds.
 ACCESSION BC036539
 VERSION BC036539.1 GI:22328116
 KEYWORDS MGC.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 2499)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (16-AUG-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK NTH-MGC Project URL: http://mgc.ncl.nih.gov
 COMMENT Contact: MGC help desk
 Email: egads-remail.nih.gov
 Tissue Procurement: Mikhail Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carnucci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Institute for Systems Biology
 http://www.systemsbio.org
 Contact: amandahsystemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
 Series: IRK Plate: 47 Row: 9 Column: 17
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Genescan gene prediction, Similarity but not identity to protein.

FEATURES
 source
 Location/Qualifiers
 1..2499
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:33863 IMAGE:5262458"
 /tissue_type="brain, hippocampus"
 /clone_lib="NTH_MGC_95"
 /lab_host="DH10B"
 /note="Vector: pBluescript"
 192..665
 /codon_start=1
 /product="similar to protein-tyrosine-phosphatase homolog DKFZP566K0524.1 - human (fragment)"
 /protein_id="AAH36539.1"
 /db_xref="GI:22328117"
 /translation="MSSPRDRAEPVNDYEGNDSEADLNREFTLPSSQENTPNSKV FENLVNSEKVKILRNPHNDYEDVEEPSESGSPSMKTANGPFRDRWSSDEDEAA GPSQALSPILSTGRTKIVSEGLDQAQIRPLIFNHEQTAKDKLTLEKNSV"

CDS
 766 a 438 c 486 g 809 t

BASE COUNT
 ORIGIN
 Alignment Scores:
 Pred. No.: 7.77e-36 Length: 2499
 Score: 369.00 Matches: 67
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 Gaps: 0

US-09-935-703-7 (1-67) x BC036539 (1-2499)

QY 1 MetGlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 670 ATGCAGCAATTATGCTTAGAATCTGACCTGCTGAGTGCATCTCGGAAAT 729
 QY 21 GlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuProPhgIlnHisGly 40
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 730 CAACCAAGCAACAGAGAAAAAACAAGATCCGAGATATCTCCATTTCAACATCATGATGA 789
 QY 41 TyrSerGlyProAsnGluArgThrPheTyrPheIleSerAsnGluGlyAlaValSer 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 790 TATAGTGGCCCAATGAGAGAACACGCTTGCGATGGTTCAACGAAGAGCATATCA 849
 QY 61 LeuLeuLeuArgTyrCysAla 67
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 850 CTTTGTTCAGATATGTGCT 870

RESULT 2
 AX354535 1914 bp DNA linear PAT 06-FEB-2002
 LOCUS AX354535
 DEFINITION Sequence 13 from Patent WO0196546.
 ACCESSION AX354535
 VERSION AX354535.1 GI:18619338
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1
 AUTHORS Au-Young, J., Baughn, M.R., Ding, L., Elliott, V.S., Gandhi, A.R., Griffin, J.A., Hatfield, A., Kearney, L., Lee, E.A., Lu, Y., Nguyen, D.B., Patterson, C., Ramkumar, J., Reddy, R., Sanjivani, M.S., Stewart, E.A., Tang, Y.T., Thornton, M., Tribouley, C.M., Yang, J., Yao, M.G. and Yue, H.
 TITLE Protein phosphatases
 JOURNAL Patent: WO 0196546-A 13 20-DEC-2001;
 INCYTE Genomics, Inc. (US)

FEATURES
 source
 Location/Qualifiers
 1..1914
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 7476861CB1"

BASE COUNT 565 a 378 c 438 g 533 t

ORIGIN
 Alignment Scores:
 Pred. No.: 5.78e-20 Length: 1914
 Score: 239.50 Matches: 66
 Percent Similarity: 26.29% Conservative: 0
 Best Local Similarity: 26.29% Mismatches: 1
 Query Match: 64.91% Indels: 185
 Gaps: 1

US-09-935-703-7 (1-67) x AX354535 (1-1914)

QY 1 MetGlnGluPheMetAlaLeuGluLeuLysAsnLeuProGlyGluPheAsnSerGlyAsn 20
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 680 ATCAGCAATTTATGCTTAGAATCTGAGATCTGCTGAGTTCAACTCTGGGAAAT 739
 QY 21 GlnProSerAsnArgGluLysAsnArgTyrArgAspIleLeuPro 35
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 740 CAACCAAGCAACAGAGAAAAAACAAGATCCGAGATATCTTCATATGATTCAACAGCC 799
 QY 35 35
 Db 800 GTTCCTCTTGAAAAAAGCAAGCATCATCATGCTATGTTATTTAGAAATGCAATTTGT 859
 QY 35 35
 Db 860 GGAGAGAGTATTTTATATCCCTACTCAAGACCACTGCTGAGCAACCATAGACTTT 919
 QY 35 35
 Db 920 TGGCAATATGCTGTGAAAAATATTTCAAAATGTTATTTGCCATGATTAACAGAGATAGAA 979
 QY 35 35


```

Db      980 GGTGAATATCAATATGCTACCTACTGGCCATTTCTGTGAAGAAAGCATTTGGAATG 1039
OY      35 ----- 35
Db      1040 AAACACTTCGCTGTATTCCTGGAGAACTACCAATACCTTCAATATTTTCATCTGCAATG 1099
OY      35 ----- 35
Db      1100 TTTCAGTTGTGAGAAATGCACAGGAACTAGTCACTGTGTAACAGTTGCAGTTCACC 1159
OY      35 ----- 35
Db      1160 AAGTGCCAGACCATGCGACTCTGCTCAGCAGATAGCTTCAATAATATATTCGTTAT 1219
OY      35 ----- 35
Db      1220 GCAAGAGAGACCACTTACAGGACCATGTTGTTCACTGCAAGTGGCCGATAGGCCGG 1279
OY      36 -----Phe 36
Db      1280 ACAGGGGTCTCTATGTGTGATGTCGTCTGTGCAATCGTAAGAACTGTTCAATTC 1339
OY      37 GlnHisIsglytyrserglyproaangluargthrhphetrphIsglySerAsnGlu 56
Db      1340 AA-CATCATGATGATATGATGCGCCCAATAGAGAAACAGTTCTGGCATGTTCAAAACGA 1398
OY      57 GlyAlaValserLeuLeuLeuArgtyrCysAla 67
Db      1399 GGAGCAGTATCACTTTTGTATGATATATGCT 1431

RESULT 3
HSM800374 2786 bp mRNA linear PRI 18-FEB-2000
LOCUS      Homo sapiens mRNA; cDNA DKFZp566K0524 (from clone DKFZp566K0524);
DEFINITION partial cds.
ACCESSION AL050040
VERSION    AL050040.1 GI:4884281
KEYWORDS
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens.
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 2786)
REFERENCE 1 (bases 1 to 2786)
AUTHORS   Ansojge,W., Winkler,U., Mewes,H.W., Gaassenhuber,J. and Wiemann,S.
TITLE     Direct Submission
JOURNAL   Submitted (15-MAY-1999) MIPS, Am Klopferspitz 18a, D-82152
            Martinsried, GERMANY
COMMENT    Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp566K0524) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://www.mips.biochem.mpg.de/proj/cDNA/.
            location/Qualifiers
FEATURES
source     1..2786
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="DKFZp566K0524"
            /tissue_type="kidney"
            /clone_id="566 (synonym: hfk42). Vector PAMP1; host
            X1-2blue; sites NotI + SalI"
            /dev_stage="fetal"
            1..1198
            /gene="DKFZp566K0524"
            <1..1198
            /gene="DKFZp566K0524"
            /note="strong similarity to protein-tyrosine-phosphatases"
            /codon_start=2

```

```

/product="hypothetical protein"
/protein_id="CAB43248.1"
/db_xref="GI:4884282"
/db_xref="SPTREMBL:Q9Y406"
/translacion="GPNMKGCGSCGTGAGAPQAVFENKYNSEKVKSLRNPHNDY
EDVEEPESGSDPSMTARGFRRDRMSSEDEEAGSQALESPLISDRKIVSEBEL
DOLAOIRPLINPHEOTAIKDCLKILEEKTAYDINOEPMALEKLKLPGEYSGNPS
NREKRNRYDILPYDSTRVPLGSKRYINARIYNGGEYFIATOGPLSTIDDFM
OMVLENSNVYATITREMEGGIIRKCYHYPIKPKLEKRRVPLENTOILOYFIIR
MEOVEKSTGTSHSVKQLEFTKMPDHPGTPASASTIKIRIRAKSHLITGPMVYHCSAG
IGRTGVFLCDVDFVFCALVYKDCSFNIMIVAQMRORSRSGNVQTRQYHFCYDYLEVLR
KLRTLD"
polya_signal 2761..2766
polya_site 2776
BASE COUNT 841 a 494 c 550 g 901 t
ORIGIN
Alignment Scores:
Pred. No.: 8,36e-19 Length: 2786
Score: 231.50 Matches: 65
Percent Similarity: 25.908 Conservative: 0
Best Local Similarity: 25.908 Mismatches: 2
Query Match: 62.74% Indels: 185
DB: 9 Gaps: 1
US-09-935-703-7 (1-67) x HSM800374 (1-2786)
OY      1 MetGlnIuphMetAlaLeuGluLeuLysAsnLeuProGlyIuphAsnSerGlyAsn 20
Db      413 ATGAGAGATTTATGCTTTAGAACTTAAGAACTCCGTGTGATGTTACTCTGGGAAT 472
OY      21 GlnProSerAsnArgGluLysAsnArgtyrArgSPILeupPro----- 35
Db      473 CAACCAAGCAACAGAAAGAAACAGATACCGAGATATTTCTCATATGATTCACACGCG 532
OY      35 ----- 35
Db      533 GTTCCCTTGAAAGAAAGCAAGCACTACATCATGCTAGTATATTAGATAGCAATTGT 592
OY      35 ----- 35
Db      593 GGAGAGACTATTTTATATCGCTACTCAAGACCACTGCTGAGCACCATAGTACTTT 652
OY      35 ----- 35
Db      653 TGGCAATGGTGTGAAATAATTAATTCATATTGCTCATGATTAACAGAGATGGAA 712
OY      35 ----- 35
Db      713 GGTGAATATCAATATGCTACATTACTGCGCCATTCTCTGAAGAGCATTGGAAATG 772
OY      35 ----- 35
Db      773 AAACACTTCGCTGATTTCTCTGAGAACTACCAATACCTTCAATATTTTCATCTCGAATG 832
OY      35 ----- 35
Db      833 TTTCAGTTGTGAGAAATGCCACGGGAATGACTCTGTAAACAGTTGCAGTTCACC 892
OY      35 ----- 35
Db      893 AAGTGCCAGACCATGCGACTCTGCTCAGAGATAGTTCATTAATAATATATTCGTTAT 952
OY      35 ----- 35
Db      953 GCAAGAGAGACCACTTACAGAGACCATGTTTTCATCTGACATGCGCGCATAGGCCGG 1012
OY      36 -----Phe 36
Db      1013 ACAGGGGTCTCTATGTGTGATGTCGTCTTGTGCCATCGTAAGAGACTGTTCAATTC 1072
OY      37 GlnHisIsglytyrserglyproaangluargthrhphetrphIsglySerAsnGlu 56
            |||||||

```

Db 1073 AA-CATCATGATATAGTACGCCCAATGAGAGAACAGCTTCTGSCATGTTCAACGAA 1131
 QY 57 GYAlaValSerLeuLeuLeuArgTyrCysAla 67
 Db 1132 GGAGCAGATATCACTTTTGTACGATATTTGCT 1164

RESULT 4
 AL603966/c
 LOCUS 103737 bp DNA linear PRI 30-JAN-2002
 DEFINITION Human DNA sequence from clone RP11-164N7 on chromosome 10, complete
 sequence.
 ACCESSION AL603966
 VERSION AL603966.9 GI:18476754
 KEYWORDS HTG:
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 103737)
 AUTHORS Bird,C.
 TITLE Direct Submission
 JOURNAL Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk
 COMMENT During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em:, EMBL; Sw:,
 SWISSPROT; Tr:, TrEMBL; Wp:, WormPEP; Information on the WormPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human
 chromosome 10, constructed by the Sanger Centre Chromosome 10
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr10
 RP11-164N7 is from the library RPCT-11.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: PBACe3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-164N7. It may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true left end of clone RP11-164N7 is at 1 in this sequence.
 The true left end of clone RP11-135E1 is at 101738 in this sequence.
 Location/Qualifiers
 1..103737
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-164N7"
 /clone_1lb="RPCT-11.1"
 BASE COUNT 31673 a 22545 c 21304 g 28215 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3.43e-08 Length: 103737
 Score: 159.00 Matches: 30
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 43.09% Indels: 0
 DB: 9 Gaps: 0

US-09-935-703-7 (1-67) x AL603966 (1-103737)
 QY 6 AAlaGluLeuLeuAsnLeuProGlyGluPheAsnSerGlyAsnGlnProSerAsnArg 25
 Db 96401 GCTTGAAGAACTTAAGAAATCTGCTGTGAGTTCAACTGTGGAAATCAACCAACGACAG 96342
 QY 26 GAluysAsnArgTyrArgAspIleuPro 35
 Db 96341 GAAAAAACAAGATACCGAGATATTTCTTCCA 96312

RESULT 5
 AL450334/c
 LOCUS 171930 bp DNA linear PRI 21-JUN-2002
 DEFINITION Human DNA sequence from clone RP11-556L1 on chromosome 10, complete
 sequence.
 ACCESSION AL450334
 VERSION AL450334.15 GI:21540028
 KEYWORDS HTG:
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 171930)
 AUTHORS Heath,P.
 TITLE Direct Submission
 JOURNAL Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk
 COMMENT During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30);
 an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em:, EMBL; Sw:,
 SWISSPROT; Tr:, TrEMBL; Wp:, WormPEP; Information on the WormPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human
 chromosome 10, constructed by the Sanger Centre Chromosome 10
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr10
 RP11-556L1 is from the library RPCT-11.2 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: PBACe3.6
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk

FEATURES
 source
 1..171930
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-556L1"
 /clone_1lb="RPCT-11.2"
 BASE COUNT 54090 a 36289 c 33202 g 48349 t
 ORIGIN

Alignment Scores:
 Pred. No.: 3.43e-08 Length: 103737
 Score: 159.00 Matches: 30
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 43.09% Indels: 0
 DB: 9 Gaps: 0

Pred. No.: 5.85e-08 Length: 171930
 Score: 159.00 Matches: 30
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 43.09% Indels: 0
 DB: 9 Gaps: 0
 US-09-935-703-7 (1-67) x AL450334 (1-171930)
 Oy 6 ALaLeuGlutLeuLeuLeuProGlyGluPheAsnSerGlyAsnGlnProSerAsnArg 25
 Db 2178 GCTTGAAGACTTGAAGATCTGCTGCTGCTGAATCAACCTCGAATCAACAGCAGCAGA 2119
 Oy 26 GtUtyAsnArgTyrTrpAspIleLeuPro 35
 Db 2118 GAAAAAACAAGATACCGAGATATCTTCCA 2089
 RESULT 6
 LOCUS AL358791 197389 bp DNA linear PRI 06-JUN-2001
 DEFINITION Human DNA sequence from clone RP11-42B19 on chromosome 10, complete
 ACCESSION AL358791
 VERSION AL358791.24 GI:1411350
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 197389)
 Clark, S.
 Direct Submission
 Submitted (02-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On May 17, 2001 this sequence version replaced g1:13751420.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw:
 SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP
 database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 10, constructed by the Sanger Centre Chromosome 10
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr10
 RP11-42B19 is from the library RPI1-11.1 constructed by the group
 of Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pBAC3.6
 This sequence is the entire insert of clone RP11-42B19.
 Location/Qualifiers
 1. 197389
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-42B19"
 /clone_1fb="RPI1-11.1"
 6. 370
 /note="match: GSS: Em:AQ116178"
 misc_feature
 26. 180

misc_feature
 76. 566
 /note="match: GSS: Em:AQ116242"
 /note="match: GSS: Em:AQ726140"
 1049. 1114
 /note="6 copies 11 mer 83% conserved"
 1192. 1246
 /note="5 copies 11 mer 87% conserved"
 1246. 1660
 /note="12 repeat: matches 2256. 2667 of consensus"
 2964. 3011
 /note="1133 repeat: matches 575. 621 of consensus"
 3170. 3580
 /note="MSTA repeat: matches 1. 426 of consensus"
 4275. 4345
 /note="MIR repeat: matches 106. 181 of consensus"
 5259. 5374
 /note="MIR repeat: matches 111. 227 of consensus"
 5414. 5454
 /note="12 repeat: matches 2657. 2697 of consensus"
 5633. 6052
 /note="TTR12 repeat: matches 1. 342 of consensus"
 6053. 6357
 /note="AluSq repeat: matches 1. 304 of consensus"
 6358. 6450
 /note="TTR12 repeat: matches 342. 440 of consensus"
 9115. 9246
 /note="12 copies 11 mer 65% conserved"
 9122. 9261
 /note="35 copies 4 mer aaga 66% conserved"
 11013. 11192
 /note="MER45 repeat: matches 1. 178 of consensus"
 11231. 11638
 /note="match: GSS: Em:AQ210526"
 11616. 12203
 /note="MER4B repeat: matches 1. 611 of consensus"
 12204. 12435
 /note="12 repeat: matches 2251. 2500 of consensus"
 12452. 12704
 /note="MIR repeat: matches 2. 262 of consensus"
 12796. 12883
 /note="12 repeat: matches 2641. 2734 of consensus"
 12939. 13288
 /note="match: GSS: Em:AZ694508"
 13577. 13719
 /note="FLAMC repeat: matches 1. 143 of consensus"
 13747. 13866
 /note="MER93 repeat: matches 110. 230 of consensus"
 13895. 14079
 /note="MER93 repeat: matches 215. 397 of consensus"
 14090. 14220
 /note="MER5B repeat: matches 25. 178 of consensus"
 14657. 14971
 /note="AluSc repeat: matches 1. 309 of consensus"
 15219. 16945
 /note="LMD repeat: matches 1905. 3500 of consensus"
 16946. 17256
 /note="AluY repeat: matches 1. 311 of consensus"
 17257. 18226
 /note="LMD repeat: matches 977. 1905 of consensus"
 18902. 19076
 /note="LMD repeat: matches 15. 174 of consensus"
 20392. 20519
 /note="MTR11 repeat: matches 6. 129 of consensus"
 20677. 20808
 /note="FLAMC repeat: matches 2. 133 of consensus"
 20833. 20868
 /note="MTR11 repeat: matches 308. 346 of consensus"
 20869. 21214
 /note="MTR1A1 repeat: matches 1. 365 of consensus"
 21217. 21621
 /note="MTR1A1-internal repeat: matches 927. 1333 of
 consensus"
 22088. 22592
 repeat_region

```

/note="LTR1A1-internal repeat: matches 160. .712 of
consensus"
repeat_region 22617. .22732
/note="LIMB3 repeat: matches 6022. .6145 of consensus"
repeat_region 22795. .22911
/note="LTR1A1-internal repeat: matches 5. .121 of
consensus"
repeat_region 22917. .23273
/note="LTR1A1 repeat: matches 1. .365 of consensus"
repeat_region 23374. .23344
/note="LTR1A1 repeat: matches 339. .410 of consensus"
repeat_region 24336. .24385
/note="LTR1F repeat: matches 268. .317 of consensus"
repeat_region 25059. .26364
/note="LIM4 repeat: matches 1064. .2552 of consensus"
repeat_region 26398. .26843
/note="LIM4 repeat: matches 2651. .3102 of consensus"
repeat_region 26879. .27248
/note="LIMB3 repeat: matches 5551. .5956 of consensus"
repeat_region 27255. .27500
/note="LTR1F repeat: matches 304. .568 of consensus"
repeat_region 27667. .27736
/note="LTR1F repeat: matches 3. .74 of consensus"
repeat_region 27759. .27927
/note="LIMB3 repeat: matches 5992. .6162 of consensus"
repeat_region 30024. .30310
/note="LIME repeat: matches 1151. .1435 of consensus"
repeat_region 31602. .31678
/note="MIR repeat: matches 79. .155 of consensus"
misc_feature 31627. .31927
/note="match: STS: Em:HSJ10C10"
repeat_region 31786. .31900
/note="MIR repeat: matches 89. .203 of consensus"
repeat_region 32440. .32738
/note="AlusX repeat: matches 1. .300 of consensus"
repeat_region 32827. .32870
/note="22 copies 2 mer lt 75% conserved"
repeat_region 34892. .35012
/note="L2 repeat: matches 2584. .2710 of consensus"
repeat_region 36225. .36549
/note="MSRD repeat: matches 21. .394 of consensus"
repeat_region 36732. .36915
/note="MIR repeat: matches 36. .234 of consensus"
repeat_region 37945. .38060
/note="MIR repeat: matches 48. .170 of consensus"
repeat_region 38422. .38524
/note="MERSA repeat: matches 9. .116 of consensus"
repeat_region 38641. .38785
/note="MIR repeat: matches 34. .192 of consensus"
repeat_region 39515. .39846
/note="LIMB3 repeat: matches 5845. .6182 of consensus"
repeat_region 40561. .41030
/note="LTR1H repeat: matches 72. .547 of consensus"
repeat_region 41704. .41789
/note="MIR repeat: matches 84. .166 of consensus"
repeat_region 41923. .42236
/note="LIME1 repeat: matches 5808. .6163 of consensus"
repeat_region 45114. .45484
/note="LTR1A1 repeat: matches 1. .365 of consensus"
repeat_region 46495. .46591
/note="LTR1J repeat: matches 43. .140 of consensus"
repeat_region 47474. .47516
/note="L2 repeat: matches 2658. .2700 of consensus"
repeat_region 48316. .48775
/note="LIMC4 repeat: matches 6443. .6923 of consensus"
repeat_region 49375. .49685
/note="AlusX repeat: matches 1. .299 of consensus"
repeat_region 49686. .49720
/note="LTR12 repeat: matches 445. .479 of consensus"
repeat_region 49691. .49724
/note="PTRS repeat: matches 173. .206 of consensus"
repeat_region 49725. .50193
/note="LTR12 repeat: matches 19. .1523 of consensus"

```

Alignment Scores:

Pred. No.:	6.76e-08	Length:	197389
Score:	159.00	Matches:	30
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	43.09%	Indels:	0
DB:	9	Gaps:	0

US-09-935-703-7 (1-67) x AL358791 (1-197389)

QY 6 AlaleuGlubeuysAsnleupProGlyGluPheAsnSerGlyAsnGlnProSerAsnArg 25

Db 178010 GCTTGAACCTTAAGAAATCTGCTGTGACTTCAACTCTGGAAATCAACGACACAGA 177951

QY 26 GtUlysAsnArgTyrrArgAspIleuPro 35

Db 177950 GAAAAAATAGATACCGAGATATTCTCCA 177921

RESULT 7

AL672108/c 65229 bp DNA linear PRI 20-AUG-2002

LOCUS Human DNA sequence from clone RP11-142117 on chromosome 10,

DEFINITION complete sequence.

ACCESSION AL672108

VERSION AL672108.7 GI:22449768

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 65229)

TITLE Direct Submission

JOURNAL Submitted (20-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquerry@sanger.ac.uk

On Aug 22, 2002 this sequence version replaced gi:22415876.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr10

RP11-142117 is from the library RGC1-11.1 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACs3.6.

Location/Qualifiers

1. .65229

/organism="Homo sapiens"

FEATURES

source

COMMENT

/db.xref="taxon:9606"
 /chromosome="10"
 /clone="RPI-142117"
 /clone_lib="RPI-11.1"
 BASE COUNT 19565 a 14038 c 1382 g 18244 t
 ORIGIN

Alignment Scores:

Pred. No.: 8 24e-06 Length: 65229
 Score: 138.00 Matches: 23
 Percent Similarity: 83.87% Conservative: 3
 Best Local Similarity: 74.19% Mismatches: 5
 Query Match: 37.40% Indels: 0
 Gaps: 0

US-09-935-703-7 (1-67) x AL672108 (1-65229)

OY 30 TYRARGASPILEUPEPHOGH1NH1SH1SGLYTYRSEGLYPROASNGIUAATGThr 49
 Db 33435 TACTGTCACTTATACAGTCAACATCATGTGATAGTGGCCCAATGAGACACACG 33376
 OY 50 PHEPHTISGLYSERASNGIUALAValSer 60
 Db 33375 TTCTGCATGTTTCAACACGAGTAGCTTCA 33343

RESULT 8
 LOCUS D64141 3090 bp mRNA linear ROD 13-FEB-1999
 DEFINITION D64141 protein-tyrosine-phosphatase, complete cds.
 ACCESSION D64141.1 GI:2665457
 VERSION D64141.1 GI:2665457
 KEYWORDS protein-tyrosine-phosphatase.
 SOURCE Mus musculus testis cDNA to mRNA.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS 1 (sites)
 TITLE Ohsugi,M., Kuramochi,S., Matsuda,S. and Yamamoto,T.
 JOURNAL Molecular cloning and characterization of a novel cytoplasmic
 MEDLINE protein-tyrosine phosphatase that is specifically expressed in
 98070510 spermatocytes
 J 1 Biol. Chem. 272 (52), 33092-33099 (1997)
 2 (bases 1 to 3090)
 REFERENCE Ohsugi,M.
 AUTHORS Direct Submission
 JOURNAL Submitted (15-SEP-1995) Minho Ohsugi, The University of Tokyo,
 Department of Oncology, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108,
 Japan (E-mail: mohnsugi@ims.u-tokyo.ac.jp, Tel:03-5449-5305,
 Fax:03-5449-5413)

FEATURES

source
 1..3090
 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /tissue_type="testis"
 218..1498
 /EC_number="3.1.3.48"
 /codon_start=1
 /product="protein-tyrosine-phosphatase"
 /protein_id="BA023761.1"
 /db_xref="GI:2665458"
 /translation="MSPPRKVGKTRGNDDEEGNSGNLNLNSLPSSSQKPTPKPI
 FGNKMSENVKPSHHLSPDYELVPEPLESDTDEVDVDSRLRNWNSMDSETA
 GSKTVPVLSGSRSLKSDTSEYSEKELTOLAOTPLIFNSASARDNCNTLOK
 EELDIIREPLELDMTLPPDNNGTCONRKRRROLIPSTVPIPLAKNDYINAS
 YIRVNHSEEFYATAGSPUPETEDFQNLNENCNVTAMITREIECGVICSTWP
 ISLKEPLEFHFVSYLETFHTVVFQIVKSKSTGSOVKHLQFTKMPDHTPA
 SADFFIVRVKRSHTITPLVHCSAGVGVICDVVFSALTEKNSEFDINIVT
 QMRKRCGMIOIKEDYOFCYEIVLEVLONLALTY"

CDS

BASE COUNT 904 a 603 c 647 g 936 t
 ORIGIN

Alignment Scores:

Pred. No.: 2.35e-05 Length: 3090
 Score: 123.00 Matches: 26
 Percent Similarity: 66.00% Conservative: 7
 Best Local Similarity: 52.00% Mismatches: 16
 Query Match: 33.33% Indels: 1
 Gaps: 0

US-09-935-703-7 (1-67) x D64141 (1-3090)

OY 1 METGNGIUPHEMETALALEUGLEULYSASNLPEUGLYGLUPHASNSEGLYasn 20
 Db 713 ATCCGTGAGTTTGTGAGTTGACAAATGACTGCTGCATGACTCAATTCGGGAAT 772
 OY 21 GINPROSEANATGULYASNAATGYTATGASPILEUPEPHOGH1NH1SH1SG 40
 Db 773 AACTACAGACAGATAGATAGACAGATATCTTCATATGATTCACACACT 832
 OY 40 YTYRSEGLYPROASNGIUAATGThr 49
 Db 833 GTTCCTTTGGAAAAACACAGACTACA 860

RESULT 9
 LOCUS RATPTPEB 892 bp mRNA linear ROD 05-FEB-1999
 DEFINITION Rat mRNA for protein tyrosine phosphatase epsilon M, partial cds.
 ACCESSION D78613
 VERSION D78613.1 GI:1228943
 KEYWORDS protein tyrosine phosphatase epsilon M; PTP epsilon M.
 SOURCE Rattus norvegicus Spleen cDNA to mRNA.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 AUTHORS 1 (bases 1 to 892)
 TITLE Nakamura,K., Mizuno,Y. and Kikuchi,K.
 JOURNAL Molecular cloning of a novel cytoplasmic protein tyrosine
 MEDLINE phosphatase ptp epsilon
 96158952 Biochem. Biophys. Res. Commun. 218 (3), 726-732 (1996)
 2 (bases 1 to 892)
 REFERENCE Nakamura,K., Mizuno,Y. and Kikuchi,K.
 AUTHORS Transmembrane and cytoplasmic protein tyrosine phosphatases PTP
 TITLE epsilon generated from a single gene
 JOURNAL unpublished

REFERENCE
 AUTHORS 3 (bases 1 to 892)
 TITLE Nakamura,K.
 JOURNAL Direct Submission

Submitted (07-DEC-1995) Koji Nakamura, Hokkaido University,
 Institute of Immunological Science, Kita-15, Nishi-7, Kita-ku,
 Sapporo, Hokkaido 060, Japan (Tel:011-707-6839, Fax:011-707-6839)
 Location/Qualifiers

FEATURES

source
 1..892
 Location/Qualifiers
 /organism="Rattus norvegicus"
 /db_xref="taxon:10116"
 /tissue_type="Spleen"
 1..>892
 /codon_start=1
 /product="protein tyrosine phosphatase epsilon M"
 /protein_id="BA011433.1"
 /db_xref="GI:1228944"
 /translation="MEPLCPILLASFSLSIATAGQNDTTPTESNMTSTTAGPPDPT
 SOLPLWLLPILLLEFLAAYPRFRKORAKAYVNSDKKPKILBEOQKRWLMS
 RSSGGRKTPPIYVEHEEETRNSAUDCKRFEENSLBSGHTGTGTELANKEENRE
 KNYFNTLPDHCRTVILSOLDGIPCSQYINASYIDGKRNKPFIAAGPKQETVNDPW
 RANWEDRSATIVM/TNLSKERKECYOWPDGCGTYGNIRVCVEDCVVLVDYTIKRF
 CIPQLPDSCKAPRLVSQHLF"

CDS

sig_peptide 1..54
 BASE COUNT 245 a 240 c 228 g 179 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.115 Length: 892
 Score: 88.50 Matches: 24

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 09:03:37 ; Search time 2414 Seconds
(without alignments)
12290.858 Million cell updates/sec

Title: US-09-935-703-2
Perfect score: 1832
Sequence: 1 ggcacgagatgatacatc.....aaataaataattactgtg 1832

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthm:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estcpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	723	39.5	775	12	BG220714 RST40503
2	714	39.0	779	14	BQ008756 UI-H-E10-
3	613.6	33.5	980	12	BG121935 602351372
4	572	31.2	607	14	BQ787020 1149a08.y
5	554.8	30.3	575	9	AA443172
6	511.8	27.9	535	14	BQ786761 1149a08.x

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	TITLE	JOURNAL	MEDLINE	COMMENT
1	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
2	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
3	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
4	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
5	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
6	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
7	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
8	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
9	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
10	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
11	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
12	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
13	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
14	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
15	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
16	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
17	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
18	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
19	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
20	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
21	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
22	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
23	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
24	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
25	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
26	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
27	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
28	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
29	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
30	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
31	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
32	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
33	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
34	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
35	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
36	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
37	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
38	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
39	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
40	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
41	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
42	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
43	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
44	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				
45	BG220714	RST40503	Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.	BG220714	GI:13746735	human.	Homo sapiens				

ALIGNMENTS

AM241948 xn77c04.x
BG655441 lb34e04.y
BG655503 lb34e04.x
AM868331 MR1-SN006
D62174 HMD243G02B
B160524 603201139
AV253789 AV253789
D79354 HMD231F08B
AK017003 Mus muscu
BH294921 CH230-44K
AV276637 AV276637
AV2744221 AV2744221
B173722 B173722
BB018630 BB018630
AV267359 AV267359
AV269368 AV269368
AV277005 AV277005
AL063921 Drosoph11
AV256976 AV256976
BB427144 BB427144
AV275486 AV275486
BB017764 BB017764
BB017796 BB017796
AL069440 Drosoph11
AL069706 Drosoph11
AQ752725 HS-5568_B
AV253804 AV253804
AL069706 Drosoph11
AL064091 Drosoph11
AV255021 AV255021
AL106171 Drosoph11
AV254458 AV254458
AV262290 AV262290
BG309087 HVSMEC000
AL176843 Tetradon
AL106940 Drosoph11
AL106940 Drosoph11
AL298972 Tetradon
AL071865 Drosoph11

FEATURES

775 bp mRNA linear EST 21-APR-2001
RST40503 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
BG220714.1 GI:13746735
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 775)
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., Mcelligott,K., Booser,S., Mays,R., Smith
E., Veloso,N., Kilka,A., Hess,J., Colthen,K., Lo,K., Offenbacher
J., Danzig,J., and Ducar,M.
Creation of genome-wide protein expression libraries using random
activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
21227151
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9800
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 549.
Location/Qualifiers

Source

```

source
1..775
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/cell_line="HT1080"
/note="See 'Creation of Genome-Wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT
213 a 144 c 164 g 254 t
ORIGIN

```

Query Match	39.5%	Score 723	DB 12	Length 775
Best Local Similarity	99.0%	Pred. No. 2,2e-127		
Matches 759; Conservative	0	Mismatches 5	Indels 3	Gaps 3

QY	129	ACATCATGGATATAGTGGCCCAATATAGAACACAGTCTGGCATGGTTCATCAACGAAG	188
Db	9	ACATCATGGATATAGTGGCCCAATATAGAG-ACAAAGTCTGGCATGGTTCATCAACGAAG	67
QY	189	AGCATATATACATTGTTGTTACGATATTTGCTCTGGAAGTCTTCGGAACTCTGACTTGG	248
Db	68	AGCATATATACATTGTTGTTACGATATTTGCTCTGGAAGTCTTCGGAACTCTGACTTGG	127
QY	249	ATTAGAAGAGCTTCGTGGTCCGTCTACCTTGAATATCCAAAGTGGTTTGCACCTCTCA	308
Db	128	ATTAGAAGAGCTTCGTGGTCCGTCTACCTTGAATATCCAAAGTGGTTTGCACCTCTCA	187
QY	309	TAAAGACATGTTTGCACTGTGCTCAAGGCTTGGCTATGACATACATCGCTTCTGG	368
Db	188	TAAAGACATGTTTGCACTGTGCTCAAGGCTTGGCTATGACATACATCGCTTCTGG	247
QY	369	TTTATCAGTTATTTTCTTCTTAAAGCTCCCTGAAGGCCAATATCATTTGGCTGGGT	428
Db	248	TTTATCAGTTATTTTCTTCTTAAAGCTCCCTGAAGGCCAATATCATTTGGCTGGGT	307
QY	429	GATCAGTGTATCTTATGATCTTGCTAGGCAATATCAAAATACCTTCCACATTTTCCA	488
Db	308	GATCAGTGTATCTTATGATCTTGCTAGGCAATATCAAAATACCTTCCACATTTTCCA	367
QY	489	GTGAACAGATGTTACATPAAACGATTCGACGCTTGGCTATTGGTTGAAGGATTACACA	548
Db	368	GTGAACAGATGTTACATPAAACGATTCGACGCTTGGCTATTGGTTGAAGGATTACACA	427
QY	549	GCCCAATPAAAGATTTAAATATATATTCATTAAAGATTTTATTTGAAAGGTGGCTGAGAG	608
Db	428	GCCCAATPAAAGATTTAAATATATATTCATTAAAGATTTTATTTGAAAGGTGGCTGAGAG	487
QY	609	AGCTGAGATTTCCAGGACTTGTAAAGTCTTATCTCTGGAGAACATPAAAGCCATATATC	668
Db	488	AGCTGAGATTTCCAGGACTTGTAAAGTCTTATCTCTGGAGAACATPAAAGCCATATATC	547
QY	669	ATGACCTCTCCAGGACTTTTAAAGACAGATGTCATTCATGTTCTTTAGCTAGAGCCGTG	728
Db	548	ATGACCTCTCCAGGACTTTTAAAGACAGATGTCATTCATGTTCTTTAGCTAGAGCCGTG	607
QY	729	TACTTTTGTGCGATTGGAATTAACCCAGTTTAAAAAGAGTCCATTAGAGGAGGAGCTAAC	788
Db	608	TACTTTTGTGCGATTGGAATTAACCCAGTTTAAAAAGAGTCCATTAGAGGAGGAGCTAAC	667
QY	789	TTTGGACACAATTTGGCTTCATTTCTACATTTTCATPACTGC-TGCCTTCTACAGCTG	847
Db	668	TTTGGACACAATTTGGCTTCATTTCTACATTTTCATPACTGCCTTCTCTACAGCTG	727
QY	848	CTAAGACCAAGACTGTGGTCTGGGAAGCAATTATGATAGGGAGA	894
Db	728	CTAAGACCAAGACTGTGGTCTGGGAAGCA-TTATGATAGGGAGAA	773

LOCUS	BQ008756	779 bp	mRNA		linear	EST 26-MAR-2002
DEFINITION	UI-H-EtIO-8yL-d-17-0-UI.s1 NCI_CGAP_EtIO Homo sapiens CDNA clone					
ACCESSION	IMAGE:5840224 3', mRNA sequence.					
VERSION	BQ008756					
KEYWORDS	BQ008756.1	GI:19733657				
SOURCE	EST.					
ORGANISM	human.					
	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
REFERENCE	1 (bases 1 to 779)					
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .					
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index					
JOURNAL	Unpublished (1997)					
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs@remail.nih.gov Tissue Procurement: Dr. Jose Mercuende cDNA Library preparation: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov The following repetitive elements were found in this cDNA sequence: 1-65, >AT-rich#low_complexity (matched complement) Seq primer: M13 FORWARD POLYA=yes Location/Qualifiers 1..779					
FEATURES						
SOURCE						

Query Match	Best Local Similarity	Score 714;	DB 14;	Length 779;
Matches 751;	Conservative 0;	Mismatches 10;	Indels 4;	Gaps 3
1068	TTAGCTACTAATTAAGTTTGATATAGCATTTATTAATTAATTCAGATCATTCATGTA	1122		
DB	TTAGCTACTAATTAAG-TTGCATATAGCATTTAT-ATATACCTGAGATCATTCAT-ATA	724		
1128	TCCTCATGTTTGATGTTATTTTCAAACTAAGATCTATGATAGTTTTTTTCCACAGTTC	1187		
DB	TCCTCATGTTTGATGTTATTTTCAAACTAAGATCTATGATAGTTTTTTTCCACAGTTC	664		
1188	CATTAAATCATTATTTCTTACTTCTACCTCTGTTGAACATTAGAAATCGATG	1247		
DB	CATTAAATCATTATTTCTTACTTCTACCTCTGTTGAACATTAGAAATCGATG	604		
1248	TGGGAACCAATTTTGGAAACCGATTCATAGTCATGAAATGGAATCTCCATATTC	1307		

Db 603 TGGGAACCAATTGTGAAAAACAGATTCATGATGAAATGGAACCTTCATATCT 544
 QY 1308 GTTTTGAAGAAGATGCGCATTTATTAAGATTTATTTAGACCTTGCCCTGACA 1367
 Db 543 GTTTTGAAGAAGATGCGCATTTATTAAGATTTATTTAGACCTTGCCCTGACA 484
 QY 1368 ATTAATAGTATTTTGTGACAAAGAGTCTGTGACAGCTATTAATTAAGCTA 1427
 Db 483 ATTAATAGTATTTTGTGACAAAGAGTCTGTGACAGCTATTAATTAAGCTA 424
 QY 1428 TAAACATATGATGATGCTTTGTACAGTTAACTCAATGAGATCGAATATCTATG 1487
 Db 423 TAAACATATGATGATGCTTTGTACAGTTAACTCAATGAGATCGAATATCTATG 364
 QY 1488 TATTGAAAAATTTTAAATCAATCTATTAATCTTAATTAAGGCTTATTTGT 1547
 Db 363 TATTGAAAAATTTTAAATCAATCTATTAATCTTAATTAAGGCTTATTTGT 304
 QY 1548 TCTTTGGCTGAATGAGTATATTTGAATGTTGTAATTAATTAATTAATTAATTA 1607
 Db 303 TCTTTGGCTGAATGAGTATATTTGAATGTTGTAATTAATTAATTAATTAATTA 244
 QY 1608 ATTAATATATGCAAAAAATATTTGATTTAAATCAATATGATGATTCCTTTACATG 1667
 Db 243 ATTAATATATGCAAAAAATATTTGATTTAAATCAATATGATGATTCCTTTACATG 184
 QY 1668 TTTCAATGATTAATTAATCTGTTTAAATTTTCAATTTTGAATTTGCTTATTTGAT 1727
 Db 183 TTTCAATGATTAATTAATCTGTTTAAATTTTCAATTTTGAATTTGCTTATTTGAT 124
 QY 1728 CTCTAATTTTATTTCTCTCTGTTAGCTAAATTAATAGCTATATGTAATTAATTTTC 1787
 Db 123 CTCTAATTTTATTTCTCTCTGTTAGCTAAATTAATAGCTATATGTAATTAATTTTC 64
 QY 1788 TTCAGAGAATTTCTAGCTATTTAATAAATAAATTTACTGTG 1832
 Db 63 TTCAGAGAATTTCTAGCTATTTAATAAATAAATTTACTGTG 19

RESULT 3
 LOCUS BGI21935 980 bp mRNA linear EST 30-JAN-2001
 DEFINITION 602351372P1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4446208 5',
 mRNA sequence.
 ACCESSION BGI21935
 VERSION BGI21935.1 GI:12615444
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM10224 row: p column: 17
 High quality sequence stop: 679.
 Location/Qualifiers
 1. 980
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4446208"
 /clone_lib="NIH_MGC_90"
 /tissue_type="adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"
 /note="Organ: Liver; Vector: pCMV-Sport6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC library."

BASE COUNT 267 a 222 c 200 g 291 t
 ORIGIN

Query Match 33.5%; Score 613.6; DB 12; Length 980;
 Best Local Similarity 97.6%; Pred. No. 1.2e-106;
 Matches 697; Conservative 0; Mismatches 9; Indels 8; Gaps 7;

QY 358 TCGTTCTGTGTTTACAGTTTATTTCTTTCTTAAAGCTCCCTGAAGGCAATATAT 417
 Db 1 TCGTTCTGTGTTTACAGTTTATTTCTTTCTTAAAGCTCCCTGAAGGCAATATAT 60
 QY 418 TCGCTTGGGGTATCAGTGTCTTATGATCTTGTGTAAGCAATATCAAAATACCTCC 477
 Db 61 TCGCTTGGGGTATCAGTGTCTTATGATCTTGTGTAAGCAATATCAAAATACCTCC 120
 QY 478 CACATTTTCCAGTGAACAGATGTTACATTAACAGATGCGCTTGGCTATTTGGTTGAA 537
 Db 121 CACATTTTCCAGTGAACAGATGTTACATTAACAGATGCGCTTGGCTATTTGGTTGAA 180
 QY 538 GGGATTACAGAGCCCAATTAAGATTTAAATATATTAATTAATTAATTTTGAAGG 597
 Db 181 GGGATTACAGAGCCCAATTAAGATTTAAATATATTAATTAATTAATTTTGAAGG 240
 QY 598 TGGCTGAGAGAGCTGAGATTTCCAGAGCTTTGTAGTCTTATTTCTGGGAGACATAA 657
 Db 241 TGGCTGAGAGAGCTGAGATTTCCAGAGCTTTGTAGTCTTATTTCTGGGAGACATAA 300
 QY 658 GGGCAATATATCATGACCTTCTCCAGCATTTTAAAGAGATCTATTCATGTTCTTA 717
 Db 301 GGGCAATATATCATGACCTTCTCCAGCATTTTAAAGAGATCTATTCATGTTCTTA 360
 QY 718 GCTAGAGCTGACTTTTGTGCGCATTTGAATTAATTAATTAATTAATTAATTAAGTCCAGTTAG 777
 Db 361 GCTAGAGCTGACTTTTGTGCGCATTTGAATTAATTAATTAATTAATTAATTAAGTCCAGTTAG 419
 QY 778 GGTGAGCTAATCTTTGGACACAA-ATTGGCTTCCATTTCTTACATTTTCTGCTGCT 836
 Db 420 GGTGAGCTAATCTTTGGACACAAATGCTTCCATTTCTTACATTTTCTGCTGCT 479
 QY 837 TCTTACAGCTGCTAGACCAAGACCTTGTGCTGGAGACATTTCAATGATAGGAGAGC 896
 Db 480 TCTTACAGCTGCTAGACCAAGACCTTGTGCTGGAGACATTTCAATGATAGGAGAGC 539
 QY 897 TCTTCTGGTGAACAGTCCAAACTTAA-ATGATGTTTATATAGAAG-CCCAAGAGA 954
 Db 540 TCTTCTGGTGAACAGTCCCAACTTAAATATGATTTTATAGAAGCCCAAGAGA 599
 QY 955 GACTTTGGCATGCTGAGTCTTTCTTATCCACCCCTTAACACTTAATTAATTAATTA 1014
 Db 600 GACTTTGGCATGCTGAGTCTTTCTTATCCACCCCTTAACACTTAATTAATTAATTA 658
 QY 1015 CTGCTTTGTAAGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1068
 Db 659 CTGCTTTGTAAGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 709

RESULT 4
 LOCUS B0787020 607 bp mRNA linear EST 26-JUL-2002
 DEFINITION 1149a08.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6033302 5',
 mRNA sequence.
 ACCESSION B0787020
 VERSION B0787020.1 GI:21995492
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 491)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the IMAGE Consortium (infoimage.lnl.gov) for further information.
Seq primer: -400P from Glibco
High quality sequence stop: 454.
Location/Qualifiers
1. 491
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2700486"
/clone_1lb="Scars_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI-CGAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as a tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Felima Bonaldo."
BASE COUNT 201 a 68 c 49 g 173 t
ORIGIN
Query Match 26.4%; Score 484.4; DB 10; Length 491;
Best Local Similarity 99.8%; Pred. No. 4.3e-82;
Matches 485; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1347 TATGAGACTTGGCTGCTACAAATTAATAGTATTTTGGACAAGAGTCTGTGACAA 1406
DB 491 TATGAGACTTGGCTGCTACAAATTAATAGTATTTTGGACAAGAGTCTGTGACAA 432
QY 1407 GGTATACCTAATTTAAGCTATTAACATAGATGAGTGTGTGACAGTTTAACTCAA 1466
DB 431 GGTATACCTAATTTAAGCTATTAACATAGATGAGTGTGTGACAGTTTAACTCAA 372
QY 1467 TGGAGATCGAATATCTATGATGAGAAATGTTAATCAATCTAATCTTGA 1526
DB 371 TGGAGATCGAATATCTATGATGAGAAATGTTAATCAATCTAATCTTGA 312
QY 1527 TTTCTAAGAGGCTTATTTGTTCTTTGGCTGAGTATGATTTGGAATGTTGAATA 1586
DB 311 TTTCTAAGAGGCTTATTTGTTCTTTGGCTGAGTATGATTTGGAATGTTGAATA 252
QY 1587 TTAATATTCGATCTAATAAATTAATATGCAAAATATTTTGATTTAAATCAA 1646
DB 251 TTAATATTCGATCTAATAAATTAATATGCAAAATATTTTGATTTAAATCAA 192
QY 1647 TAGATGATTCGTGTACATGTTCAATATGATATGATGATGATTTGATTTGATA 1706
DB 191 TAGATGATTCGTGTACATGTTCAATATGATATGATGATGATTTGATTTGATA 132
QY 1707 ATGGCCCTTAATATTTGATCTAATATTTTCTCTGCTGATCTGTAATAATAG 1766
DB 131 ATGGCCCTTAATATTTGATCTAATATTTTCTCTGCTGATCTGTAATAATAG 72
QY 1767 CTATATGATTAACAATTTCTTCGAGAGAAATTCATGCTATATTAATAAATATTT 1826
DB 71 CTATATGATTAACAATTTCTTCGAGAGAAATTCATGCTATATTAATAAATATTT 12

QY 1827 ACTGTG 1832
DB 11 ACTGTG 6
RESULT 8
BG655441
LOCUS BG655441
DEFINITION 499 bp mRNA linear EST 05-JUL-2001
ACCESSION 1b34e04.y1 HR85 1stet Homo sapiens cDNA 5', mRNA sequence.
VERSION BG655441.1 GI:13792850
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 499)
Lemishka, I., Scearce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blais, A., Schmitt, A., Reising, B., Ralder, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, V., Williams, T., Jackson, Y., and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8537
Email: dmelton@bioh.p.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -400P from Glibco
High quality sequence stop: 446.
Location/Qualifiers
1. 499
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="HR85 1stet"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site.1: NotI; Site.2: XhoI; cDNA made by oligo-dT priming. Size selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permut Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@im.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."
BASE COUNT 155 a 66 c 86 g 192 t
ORIGIN
Query Match 25.9%; Score 474; DB 12; Length 499;
Best Local Similarity 99.8%; Pred. No. 4.1e-80;
Matches 485; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1105 ATAAATCTGAGTCATTCATGTCATCTCATGTTGATGATTTTCAACATAGATCTA 1164
DB 14 ATAAATCTGAGTCATTCATGTCATCTCATGTTGATGATTTTCAACATAGATCTA 73
QY 1165 TGAATG-TTTTTTCCAGAGTTCATTAATTCATTTATTTCTTACTTCTCACTCT 1223
DB 74 TGAATGTTTTTTTTCCAGAGTTCATTAATTCATTTATTTCTTACTTCTCACTCT 133
QY 1224 GTTGAACATTTTGAACATGATTTGGACCCCAATTTTGGAAACAGATTCATAGCA 1283
DB 134 GTTGAACATTTTGAACATGATTTGGACCCCAATTTTGGAAACAGATTCATAGCA 193

QY 1284 TGAAGATGAAATTCATATTCGTTTGAAGAGATGCGCATTTACAGTAATT 1343
 Db 194 TGAAGATGAAATTCATATTCGTTTGAAGAGATGCGCATTTACAGTAATT 253
 QY 1344 TATATAGAGACTTGGCTGCTACATATTAATAGATATTTTGAAGAGAGAGTCTGTGA 1403
 Db 254 TATATAGAGACTTGGCTGCTACATATTAATAGATATTTTGAAGAGAGAGTCTGTGA 313
 QY 1404 CAAGCTATACCTATTAATAGCTATTAAGATATTAAGATATTAAGATATTAAGTCTGTGA 1463
 Db 314 CAAGCTATACCTATTAATAGCTATTAAGATATTAAGATATTAAGATATTAAGTCTGTGA 373
 QY 1464 CAATGAGATCAGATATTAATATTAATAGATATTAAGATATTAAGATATTAAGTCTGTGA 1523
 Db 374 CAATGAGATCAGATATTAATATTAATAGATATTAAGATATTAAGATATTAAGTCTGTGA 433
 QY 1524 GAATTTCTAAGAGCTTATTTTCTTTTGGCTGAATGAGATATTAAGATATTAAGTCTGTGA 1583
 Db 434 GAATTTCTAAGAGCTTATTTTCTTTTGGCTGAATGAGATATTAAGATATTAAGTCTGTGA 493
 QY 1584 TAAATTA 1589
 Db 494 TAAATTA 499

RESULT 9
 BG655503/c 396 bp mRNA linear EST 05-JUL-2001
 LOCUS 1B34E04.x1 HR85 1slet Homo sapiens cDNA 3', mRNA sequence.
 DEFINITION BG655503
 VERSION BG655503.1 GI:13792912
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 396)
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
 Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Hillier, L., Marra, M., Pape, D., Wyllie, T., Martin, O., Blistain, A.,
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas,
 M., Gibbons, M., McCann, R., Cole, R., Tsagarashvili, R., Williams, T.,
 Jackson, Y., and Bowers, Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)
 Other ESTs: 1B34E04.y1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8537
 Email: dmelton@biochem.harvard.edu
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Dr. Hiroshi Inoue
 (hinoue@im.wustl.edu)
 Seq primer: -400P from Gibco
 High quality sequence stop: 337.
 Location/Qualifiers
 1. 396
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HR85 1slet"
 /tissue_type="Purified pancreatic islet"
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site:1:
 NotI; Site:2: XhoI; cDNA made by Oligo-dT priming.
 Size selected on agarose gel. Average insert size ~1kb. 5'
 XhoI site was destroyed after directional cloning.
 Amplified once. Contact information: Hiroshi Inoue, MD,

Metabolism Div. (Alan Permutt Lab), Washington University
 School of Medicine, Box 8127, 660 South Euclid Ave., St.
 Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
 314-362-1916, Fax: 314-747-2692."

BASE COUNT 161 a 46 c 35 g 154 t
 ORIGIN

Query Match 19.5%; Score 357; DB 12; Length 396;
 Best Local Similarity 99.7%; Pred. No. 6, 6e-58;
 Matches 368; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1463 TCAATGAGATCAGATATTAATATTAATAGATATTAAGATATTAAGATATTAAGTCTGTGA 1522
 Db 396 TCAATGAGATCAGATATTAATATTAATAGATATTAAGATATTAAGATATTAAGTCTGTGA 337
 QY 1523 TGAATTTCTAAGAGCTTATTTTCTTTTGGCTGAATGAGATATTAAGATATTAAGTCTGTGA 1582
 Db 336 TGAATTTCTAAGAGCTTATTTTCTTTTGGCTGAATGAGATATTAAGATATTAAGTCTGTGA 277
 QY 1583 ATAATTAATTAATTCATGTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1642
 Db 276 ATAATTAATTAATTCATGTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 217
 QY 1643 CAATGAGATCAGATATTAATATTAATAGATATTAAGATATTAAGATATTAAGTCTGTGA 1702
 Db 216 CAATGAGATCAGATATTAATATTAATAGATATTAAGATATTAAGATATTAAGTCTGTGA 157
 QY 1703 GATTAATGGCTTTAATATTAATTTGATCTGATATTAATTTTCTGCTGCTGTAATAATA 1762
 Db 156 GATTAATGGCTTTAATATTAATTTGATCTGATATTAATTTTCTGCTGCTGTAATAATA 97
 QY 1763 ATAGCTATTAATTAATTAATTAATTTCTGGAAGATATTAATTAATTAATTAATTAATTAAT 1822
 Db 96 ATAGCTATTAATTAATTAATTTCTGGAAGATATTAATTAATTAATTAATTAATTAATTAAT 38
 QY 1823 ATTTACTGT 1831
 Db 37 ATTTACTGT 29

RESULT 10
 AM868331 302 bp mRNA linear EST 22-MAY-2000
 LOCUS MR1-SN0060-050500-002-f11 SN0060 Homo sapiens cDNA, mRNA sequence.
 DEFINITION AM868331
 ACCESSION AM868331.1 GI:8002363
 VERSION AM868331.1
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 302)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, N. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 200202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=st2-MR1-SN0060-050

FEATURES

Source

500-002-fil1e13-2000-05-05e14-1)
 Seq primer: puc 18 forward
 High quality sequence start: 61
 High quality sequence stop: 302.
 Location/Qualifiers
 1.302

FEATURES

source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="SN0060"
 /dev_stage="Adult"
 /note="Organ: stomach normal. Vector: puc18; Site.1: Sma1; Site.2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 82 a 68 c 59 g 93 t

ORIGIN

Query Match 15.6%; Score 286; DB 10; Length 302;
 Best Local Similarity 98.3%; Pred. No. 2e-44;
 Matches 289; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 657 AGCCCAATATCATATACCTCTTCCAGGCAATTTTAAGACAGATGCTATTCATGTTCTTT 716
 DB 7 AGCGAACTGTCATACCTCTTCCAGGCAATTTTAAGACAGATGCTATTCATGTTCTTT 66
 QY 717 AGCTAGACCTCTGATCTTTTCTGCAATTTGAATACCCAGTTAAAGAGCCAGTTA 776
 DB 67 AGCTAGACCTCTGATCTTTTCTGCAATTTGAATACCCAGTTAAAGAGCCAGTTA 126
 QY 777 GGGTGGCACTAATCTTGGACACAAATGGCTTCATTCTCTACATTTTCTACTGCTCT 836
 DB 127 GGGTGGCACTAATCTTGGACACAAATGGCTTCATTCTCTACATTTTCTACTGCTCT 186
 QY 837 TCTTACAGCTGCTATACCAAGACCTGGTGGTGGGAACATTTCTATGATAGGAGAGC 896
 DB 187 TCTTACAGCTGCTATACCAAGACCTGGTGGTGGGAACATTTCTATGATAGGAGAGC 246
 QY 897 TCTCTCTGGTGAACAGTCCAAATCTAAATATAGATGTTTATATAGAAAGCCCAAG 950
 DB 247 TCTCTCTGGTGAACAGTCCAAATCTAAATATAGATGTTTATATAGAAAGCCCAAG 300

RESULT 11

LOCUS D62174 260 bp mRNA linear EST 29-AUG-1995
 DEFINITION HUM243G02B Clontech human aorta polyA+ mRNA (#6572) Homo sapiens
 ACCESSION D62174
 VERSION D62174.1 GI:965950
 KEYWORDS EST.

SOURCE

human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 260)
 Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takai, C., Shin, S. and Nakamura, Y.
 Fujiwara et al. (1995)
 Unpublished (1995)

COMMENT

Contact: Tsutomu Fujiwara
 Otsuka GEM Research Institute
 Otsuka Pharmaceutical Co., Ltd
 463-10 Kagasuno Kawachi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035.

FEATURES

Location/Qualifiers
 1.260
 /organism="Homo sapiens"

/db_xref="taxon:9606"
 /clone_1lb="Clontech human aorta polyA+ mRNA (#6572)"
 /note="Male adult, hematopoietic tissue, stem cell"

BASE COUNT

87 a 21 c 34 g 109 t

ORIGIN

Query Match 13.9%; Score 255.2; DB 14; Length 260;
 Best Local Similarity 96.5%; Pred. No. 1.4e-38;
 Matches 251; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

QY 1444 AGTGTGTACAGTTTAACTCAATGAGATGAGATATTCATGATGAGAAATGTTT 1503
 DB 1 AGTGTGTACAGTTTAACTCAATGAGATGAGATATTCATGATGAGAAATGTTT 60
 QY 1504 AATATCAATCTATTAATCTTGAATTTCTAAGAGCTTATTTGTTCTTTGGCTGAATCA 1563
 DB 61 AATATCAATCTATTAATCTTGAATTTCTAAGAGCTTATTTGTTCTTTGGCTGAATCA 120
 QY 1564 GATATTTGAATTTGTTGAATTAATTAATTTCTATGTAATAATTAATATGCAAA 1623
 DB 121 GATATTTGAATTTGTTGAATTAATTAATTTCTATGTAATAATTAATATGCAAA 180
 QY 1624 AATATTTGATGTTTAAATCAATAGATGATCTGTTTACATTTGTTCAATATGATTA 1683
 DB 181 AATATTTGATGTTTAAATCAATAGATGATCTGTTTACATTTGTTCAATATGATTA 240
 QY 1684 TCTGTGTTAATTTCAATTTG 1703
 DB 241 TCTGTGTTAATTTCAATTTG 260

RESULT 12

LOCUS B1460524 571 bp mRNA linear EST 21-AUG-2001
 DEFINITION 603201139F1 NIH-MGC_97 Homo sapiens cDNA clone IMAGE:5267145 5',
 mRNA sequence.

ACCESSION B1460524
 VERSION B1460524.1 GI:15251180
 KEYWORDS EST.

SOURCE

human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 571)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strauberg, Ph.D.
 Email: cgabs-remail.nih.gov

COMMENT

Tissue Procurement: Miklos Palcovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
 Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMNL at:
 http://image.llnl.gov
 Plate: LMNL1673 row: n column: 10
 High quality sequence stop: 569.

FEATURES

Location/Qualifiers
 1.571

SOURCE

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="IMAGE:5267145"
 /clone_1lb="NIH-MGC_97"
 /lab_host="DH10B"
 /note="Organ: testis; Vector: pBluescript (modified pBluescript KS+); Site.1: BamHI; Site.2: SalI; XhoI (gtcagag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 11:04:12 ; Search time 262 seconds
(Without alignments)
9786.078 Million cell updates/sec

Title: US-09-935-703-2

Perfect score: 1832
Sequence: 1 ggcacgaggtatgatacat.....aaataaataattactctg 1832

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 870385 seqs, 699768693 residues

Total number of hits satisfying chosen parameters: 1740770

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEM_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEM_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEM_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEM_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEM_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEM_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEM_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	360.6	19.7	381	9	US-10-076-622-208 Sequence 208, App
2	360.6	19.7	381	10	US-09-604-287A-208 Sequence 208, App
3	360.6	19.7	381	10	US-09-339-338-208 Sequence 208, App
4	360.6	19.7	381	12	US-10-007-805-208 Sequence 208, App
5	71.8	3.9	15732	9	US-10-239-676-95 Sequence 95, App1
6	70	3.8	5689	9	US-10-239-676-90 Sequence 90, App1
7	67.8	3.7	11047	9	US-10-239-676-187 Sequence 187, App
8	67.6	3.7	6030	9	US-10-239-676-164 Sequence 164, App
9	65.4	3.6	7823	9	US-10-239-676-197 Sequence 197, App
10	65.2	3.6	4985	9	US-10-094-240-10 Sequence 10, App1
11	64	3.5	502	9	US-10-198-846-8860 Sequence 8860, App
12	64	3.5	5979	9	US-10-239-676-18 Sequence 18, App1
13	61.6	3.4	6620	9	US-10-239-676-195 Sequence 195, App
14	61.2	3.3	9539	9	US-10-239-676-51 Sequence 51, App1
15	61.2	3.3	11047	9	US-10-239-676-188 Sequence 188, App
16	60.6	3.3	3007	9	US-10-239-676-219 Sequence 219, App
17	60.6	3.3	9515	9	US-10-239-676-160 Sequence 160, App1
18	60	3.3	15732	9	US-10-239-676-96 Sequence 96, App1
19	59.6	3.3	6337	9	US-10-239-676-151 Sequence 151, App

20	59.2	3.2	8588	9	US-10-239-676-177 Sequence 177, App
21	58.8	3.2	169139	9	US-10-067-514-1 Sequence 1, App1
22	58.2	3.2	5979	9	US-10-239-676-18 Sequence 18, App1
23	58	3.2	4985	9	US-10-094-240-10 Sequence 10, App1
24	57.8	3.2	6306	9	US-10-239-676-129 Sequence 129, App1
25	57.4	3.1	11260	9	US-10-239-676-190 Sequence 19, App1
26	57.2	3.1	5610	9	US-10-239-676-170 Sequence 170, App1
27	57.2	3.1	11812	9	US-10-239-676-210 Sequence 210, App
28	57.2	3.1	17421	9	US-10-239-676-53 Sequence 53, App1
29	56.8	3.1	431	10	US-09-960-352-5558 Sequence 5558, App
30	56.2	3.1	516	10	US-09-960-352-5785 Sequence 5785, App
31	56.2	3.1	8866	9	US-10-239-676-140 Sequence 140, App
32	56	3.1	416	10	US-09-960-352-4584 Sequence 4584, App
33	55.8	3.0	9539	9	US-10-239-676-52 Sequence 52, App1
34	55.6	3.0	960	9	US-10-198-846-6381 Sequence 6381, App1
35	55.4	3.0	377	10	US-09-960-352-7419 Sequence 7419, App
36	55.2	3.0	17848	9	US-10-239-676-28 Sequence 28, App1
37	55	3.0	480	10	US-09-960-352-5301 Sequence 5301, App
38	55	3.0	2000	9	US-09-938-842A-5191 Sequence 5191, App
39	55	3.0	5689	9	US-10-239-676-89 Sequence 89, App1
40	55	3.0	6053	9	US-10-239-676-76 Sequence 76, App1
41	55	3.0	11836	9	US-10-239-676-102 Sequence 102, App
42	55	3.0	14649	9	US-10-239-676-122 Sequence 122, App
43	54.8	3.0	6298	9	US-10-239-676-64 Sequence 64, App1
44	54.8	3.0	53332	9	US-10-224-562-3 Sequence 3, App1
45	54.8	3.0	53332	10	US-09-801-861-3 Sequence 3, App1

ALIGNMENTS

RESULT 1
US-10-076-622-208

Sequence 208, Application US/10076622
Publication No. US20030023036A1

GENERAL INFORMATION:

APPLICANT: Houghton, Raymond L.

APPLICANT: Sleath, Paul R.

APPLICANT: Persing, David H.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

FILE REFERENCE: 210121.470C11

CURRENT APPLICATION NUMBER: US/10/076,622

CURRENT FILING DATE: 2002-02-13

NUMBER OF SEQ ID NOS: 627

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 208

LENGTH: 381

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc.feature

LOCATION: 10, 27, 37, 46, 75, 95, 102, 137, 143, 202, 234, 278, 310,

LOCATION: 351

OTHER INFORMATION: n = A,T,C or G

US-10-076-622-208

Query Match
Best Local Similarity 95.3%; Pred. No. 8.7e-60;
Matches 363; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 949 AGAGAGACTTTGGCAGTCGAGTCTTTCTTCCATCCACCCCTACACTTACATATTA 1008

DB 1 AGAGAGAGATNTTGGCAGTCGAGTCTTTCTTCCATCCACCCCTACACTTACATATTA 60

QY 1009 CTTAGTCTGCTTTGTTAAAGCAAGATTAACCTTAAGTCTTACCTTGGCCTT 1068

DB 61 CTTAGTCTGCTTTGTTAAAGCAAGATTAACCTTAAGTCTTACCTTGGCCTT 120

QY 1069 TAGCTAATTAAGATTTGATATAGCATTTATATATATCTGAGTCATTCATGATAT 1128

DB 121 TAGCTAATTAAGATTTGATATAGCATTTATATATATCTGAGTCATTCATGATAT 180

QY 1129 CTCATCTGTTGATGATTTTTCACAACTAAGATCTATGATGTTTTCCTCAGAGTCC 1188
DB 181 CTCATCTGTTGATGATTTTTCACAACTAAGATCTATGATGTTTTCCTCAGAGTCC 240
QY 1189 ATTAATCATTTATTTCTTACTCTCCCTGTTGTTGTTTTCAGAACTGAGATTT 1248
DB 241 ATTAATCATTTATTTCTTACTCTCCCTGTTGTTGTTTTCAGAACTGAGATTT 300
QY 1249 GGAACCCCAATTTTGGAAAACAGATTCATGATGAAAATGGAACCTTCATATTCG 1308
DB 301 GGAACCCCAATTTTGGAAAACAGATTCATGATGAAAATGGAACCTTCATATTCG 360
QY 1309 TTTTGAAGAGATGTGGCAT 1329
DB 361 TTTTGAAGAGATGTGGCAT 381

RESULT 2

US-09-604-287A-208
Sequence 208, Application US/09604287A
Patent No. US20020064872A1
GENERAL INFORMATION:
APPLICANT: Jiang, Yugu
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Xu, Jiangchun
APPLICANT: Harlocker, Susan L.
APPLICANT: Hepler, William T.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.470C7
CURRENT APPLICATION NUMBER: US/09/604,287A
CURRENT FILING DATE: 2000-06-22
NUMBER OF SEQ ID NOS: 489
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 208
LENGTH: 381
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)...(381)
OTHER INFORMATION: n-A,T,C or G
US-09-604-287A-208

Query Match 19.7%; Score 360.6; DB 10; Length 381;
Best Local Similarity 95.3%; Pred. No. 8.7e-60;
Matches 363; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 949 AGAGAGACTTTTGGCATGCGCTGAGTCTTCTCTATCCCACTTAACATATTA 1008
DB 1 AGAGAGACTTTTGGCATGCGCTGAGTCTTCTCTATCCCACTTAACATATTA 60
QY 1009 CTTAGCTGCTTTGTTAAAGCAAGTATACCTTTAGCTGCTTACTCTTGCCCTT 1068
DB 61 CTTAGCTGCTTTGTTAAAGCAAGTATACCTTTAGCTGCTTACTCTTGCCCTT 120
QY 1069 TAGCTAACTAATAAGTTGATATAGGCATTTATATATATCTGATCATGATAT 1128
DB 121 TAGCTAACTAATAAGTTGATATAGGCATTTATATATATCTGATCATGATAT 180
QY 1129 CTCATCTGTTGATGATTTTTCACAACTAAGATCTATGATGTTTTCAGAGTCC 1188
DB 181 CTCATCTGTTGATGATTTTTCACAACTAAGATCTATGATGTTTTCAGAGTCC 240
QY 1189 ATTAATCATTTATTTCTTACTCTCCCTGTTGTTGTTTTCAGAACTGAGATTT 1248
DB 241 ATTAATCATTTATTTCTTACTCTCCCTGTTGTTGTTTTCAGAACTGAGATTT 300
QY 1249 GGAACCCCAATTTTGGAAAACAGATTCATGATGAAAATGGAACCTTCATATTCG 1308
DB 301 GGAACCCCAATTTTGGAAAACAGATTCATGATGAAAATGGAACCTTCATATTCG 360

QY 1309 TTTTGAAGAGATGTGGCAT 1329
DB 361 TTTTGAAGAGATGTGGCAT 381

RESULT 3

US-09-339-338-208
Sequence 208, Application US/09339338A
Patent No. US20020102602A1
GENERAL INFORMATION:
APPLICANT: Yugu, Jiang
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Xu, Jiangchun
TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
FILE REFERENCE: 210121.470C2
CURRENT APPLICATION NUMBER: US/09/339,338A
CURRENT FILING DATE: 1999-06-23
NUMBER OF SEQ ID NOS: 315
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 208
LENGTH: 381
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)...(381)
OTHER INFORMATION: n-A,T,C or G
US-09-339-338-208

Query Match 19.7%; Score 360.6; DB 10; Length 381;
Best Local Similarity 95.3%; Pred. No. 8.7e-60;
Matches 363; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 949 AGAGAGACTTTTGGCATGCGCTGAGTCTTCTCTATCCCACTTAACATATTA 1008
DB 1 AGAGAGACTTTTGGCATGCGCTGAGTCTTCTCTATCCCACTTAACATATTA 60
QY 1009 CTTAGCTGCTTTGTTAAAGCAAGTATACCTTTAGCTGCTTACTCTTGCCCTT 1068
DB 61 CTTAGCTGCTTTGTTAAAGCAAGTATACCTTTAGCTGCTTACTCTTGCCCTT 120
QY 1069 TAGCTAACTAATAAGTTGATATAGGCATTTATATATATCTGATCATGATAT 1128
DB 121 TAGCTAACTAATAAGTTGATATAGGCATTTATATATATCTGATCATGATAT 180
QY 1129 CTCATCTGTTGATGATTTTTCACAACTAAGATCTATGATGTTTTCAGAGTCC 1188
DB 181 CTCATCTGTTGATGATTTTTCACAACTAAGATCTATGATGTTTTCAGAGTCC 240
QY 1189 ATTAATCATTTATTTCTTACTCTCCCTGTTGTTGTTTTCAGAACTGAGATTT 1248
DB 241 ATTAATCATTTATTTCTTACTCTCCCTGTTGTTGTTTTCAGAACTGAGATTT 300
QY 1249 GGAACCCCAATTTTGGAAAACAGATTCATGATGAAAATGGAACCTTCATATTCG 1308
DB 301 GGAACCCCAATTTTGGAAAACAGATTCATGATGAAAATGGAACCTTCATATTCG 360
QY 1309 TTTTGAAGAGATGTGGCAT 1329
DB 361 TTTTGAAGAGATGTGGCAT 381

RESULT 4

US-10-007-805-208
Sequence 208, Application US/10007805
Patent No. US20020150581A1
GENERAL INFORMATION:
APPLICANT: Jiang, Yugu
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Xu, Jiangchun

APPLICANT: Harlocker, Susan L.
APPLICANT: Heppler, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick, Thomas S.
APPLICANT: McNeill, Patricia D.
APPLICANT: Durham, Margarita
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.470C10
CURRENT APPLICATION NUMBER: US/10/007,805
NUMBER OF SEQ ID NOS: 593
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 208
LENGTH: 381
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
LOCATION: 10, 27, 37, 46, 75, 95, 102, 137, 143, 202, 234, 278, 310,
LOCATION: 351
OTHER INFORMATION: n = A,T,C or G
US-10-007-805-208

Query Match 19.7%; Score 360.6; DB 12; Length 381;
Best Local Similarity 95.3%; Pred. No. 8,7e-60;
Matches 353; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 949 AGAGGAGCATTTGGCATGCGTTCAGTTCCTTCCATCCCACTTAACATTA 1008
DB 1 AAGGAGATNTTGGCATGCGTTCAGTTCCTTCCATCCCACTTAACATTA 60
QY 1009 CTTAGTCTCTTTGTTAAAGCAAGTATTTACCTTAAGTCCCTTACCTTTGGCCTT 1068
DB 61 CTTAGTCTCTTTGTTAAAGCAAGTATTTACCTTAAGTCCCTTACCTTTGGCCTT 120
QY 1069 TAGCTACTAATTAAGTGTGATATAGCATTAATTAATTTGAGTCAATTCATG 1128
DB 121 TAGCTACTAATTAAGTGTGATATAGCATTAATTAATTTGAGTCAATTCATG 180
QY 1129 CTCCTATGTTGATGATTTTCAACTAAGATCTATGATGATTTTTCAGAGTTC 1188
DB 181 CTCCTATGTTGATGATTTTCAACTAAGATCTATGATGATTTTTCAGAGTTC 240
QY 1189 ATTAATCATTTATTTCTTACTTTCTACCTCTGTGAAACATTTAGAACTGAT 1248
DB 241 ATTAATCATTTATTTCTTACTTTCTACCTCTGTGAAACATTTAGAACTGAT 300
QY 1249 GGAACCCCAATTTGGAAACCAAGATTCATAGCATGAAATGGAACCTTCATTT 1308
DB 301 GGAACCCCAATTTGGAAACCAAGATTCATAGCATGAAATGGAACCTTCATTT 360
QY 1309 TTTTGAAGAAGTGGGCAT 1329
DB 361 TTTTGAAGAAGTGGGCAT 381

RESULT 5
US-10-239-676-95
Sequence 95, Application US/10239676
Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239,676
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10019173.8

DE 10032529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 95
LENGTH: 15732
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-95

Query Match 3.9%; Score 71.8; DB 9; Length 15732;
Best Local Similarity 49.9%; Pred. No. 0.00065;
Matches 208; Conservative 0; Mismatches 207; Indels 2; Gaps 1;

QY 1416 AATTATAGCTTAAACATATGATATGAGTGTGTTGTACAGTTTAACTCAATGAGATCA 1475
DB 4351 AGTAATTATGAAATTAATTAATTAAGTGAAGTATGATTAATTAATTAAGAGATTA 4410
QY 1476 GAATATTCATGATTTGAGAAATGTTTATATCATCATATAATCTTGAAATTCAGAGA 1535
DB 4411 TTAGATGAAATTTTAAATTAATTTGAAATGAGATTAATTAATTAATTAATTAATTT 4470
QY 1536 GCCTTATTTTCTTTGCTTTGGCTGAATGAGTATATTTGAATGGTGAATTAATTAAT 1595
DB 4471 G--TTTGTGTAAGTTTGTGTTGTTGTTTACTTATTAATTTGGTGAATATGAGTAT 4528
QY 1596 CTGATTTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1655
DB 4529 ATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4588
QY 1656 CTGTTACATTTGTCATATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 1715
DB 4589 TGGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4648
QY 1716 TAAATTTGTAATCTTAATTTTATTTTCTCTGTTACTGTAATTAATTAATTAAT 1775
DB 4649 GAGGAGAAATTTTGTGTTGTTTATTTTATTAATTAATTAATTAATTAATTAAT 4708
QY 1776 ATTAACAATTTCTTACAGAAATTTCTATGCTATTTTAATTAATTAATTAATTA 1832
DB 4709 TTATCGGTTTAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4765

RESULT 6
US-10-239-676-90
Sequence 90, Application US/10239676
Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239,676
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 90

LENGTH: 5689
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 NAME/KEY: unsure
 LOCATION: (1818, 1930..1931, 1934..1935, 1947..1948, 1966)
 US-10-239-676-90

Query Match 3.8%; Score 70; DB 9; Length 5689;
 Best Local Similarity 47.7%; Pred. No. 0.00094;
 Matches 285; Conservative 0; Mismatches 286; Indels 4; Gaps 2;

QY 1274 TTCAATGTCATGAAGAAAGAACTCCATATTCGTTTGAAGAAAGTGGCCATTAT 1333
 DB 1290 TTTATTTTAT 1349
 QY 1334 ACAGTAATTTTAT 1393
 DB 1350 ATATTTATTTTATATATATATATATATATATATATATATATATATATAT 1409
 QY 1394 GTTCGTGACAGCTATACCTATATATATATATATATATATATATATATATAT 1453
 DB 1410 TATATTTTAAATTAATTTTATATATATATATATATATATATATATATAT 1469
 QY 1454 CAGTTTAACTCAATGAGAG-TCAAGATATTTCTATATATATATATATATATAT 1512
 DB 1470 TTTATTTATTCGAGATTAATATATATATATATATATATATATATATATAT 1529
 QY 1513 CTATTAATCTGTAATTTCTAAGAGCTTATATATATATATATATATATATAT 1572
 DB 1530 TATATTTTAT 1589
 QY 1573 AATTGTTGAT 1632
 DB 1590 ATATTTTAT 1649
 QY 1633 GATGTTAAATCAATATAGATATCTGTTTACATTTCTCATATATATATATAT 1692
 DB 1650 TTTATTTATATGATTAATATATATATATATATATATATATATATATATAT 1709
 QY 1693 ATTTCATTTTGAATATATGCGCTTATATATATATATATATATATATATAT 1752
 DB 1710 ATT- -TTTAAATATTTTATATATATATATATATATATATATATATAT 1766
 QY 1753 CTGTAAATATATAGCTATATATATATATATATATATATATATATATATAT 1812
 DB 1767 TTTATATATTTTATATATATATATATATATATATATATATATATATAT 1826
 QY 1813 AAATATAAATATTTTA 1827
 DB 1827 TTTATTTTAAATATATA 1841

RESULT 7
 US-10-239-676-187
 Sequence 187, Application US/10239676
 Publication No. US20030082609A1
 GENERAL INFORMATION:
 APPLICANT: OLEK, Alexander
 APPLICANT: PIEPENBROCK, Christian
 TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
 FILE REFERENCE: 5013.1003
 CURRENT APPLICATION NUMBER: US/10/239, 676
 CURRENT FILING DATE: 2002-09-24
 PRIOR APPLICATION NUMBER: PCT/EP01/03968
 DE 10019058.8
 DE 10019173.8
 DE 10032529.7
 DE 10043826.1
 PRIOR FILING DATE: 2001-04-06

2000-04-06
 2000-04-07
 2000-06-30
 2000-09-01
 NUMBER OF SEQ ID NOS: 228
 SEQ ID NO 187
 LENGTH: 11047
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
 US-10-239-676-187

Query Match 3.7%; Score 67.8; DB 9; Length 11047;
 Best Local Similarity 43.5%; Pred. No. 0.0033;
 Matches 306; Conservative 0; Mismatches 397; Indels 0; Gaps 0;

QY 1087 TGATATAGCATTTAT 1146
 DB 9873 TGATATATATTTGTTATATATATATATATATATATATATATATATATAT 9932
 QY 1147 TTTTCAACTAGATCTATATATATATATATATATATATATATATATATAT 1206
 DB 9933 TATATGGGTGATATATATATATATATATATATATATATATATATATAT 9992
 QY 1207 TTTACTTCTCACCTCTGTTGAACAATTTAGAAACTGATTTGGGAACCAATTT 1266
 DB 9993 ATATTTATTTGGAAATTAAGAAATTTTAAATTTTATATATATATATAT 10052
 QY 1267 AACCAATTCATATGATCAAGAAATGAAACTTCATATCTGTTTGAAGAAATG 1326
 DB 10053 GTGTGTAGGAAATTAAGAAATTTGAAGAAATTTGTATATTTGTTAGGTT 10112
 QY 1327 CATTTTACAGTAATTTTATATATATATATATATATATATATATATATAT 1386
 DB 10113 TATTAATATTTATATATATATATATATATATATATATATATATATAT 10172
 QY 1387 ACAAGAGTTCGTCGACCAAGCTATACCTATATATATATATATATATATAT 1446
 DB 10173 GATGATTTTAAAGTAATTTAGATATATATATATATATATATATATAT 10232
 QY 1447 GTTTGACGTTTAACTCAATGAGATCAGATATATATATATATATATATAT 1506
 DB 10233 AGTTGATATATATTTTGGTTTTTTTATATATATATATATATATATAT 10292
 QY 1507 ATCAATCTAATATCTGTAATTTCTAAGAGCTTATATATATATATATATAT 1566
 DB 10293 TTTTAT 10352
 QY 1567 TATTTGAATGCTTGAATATATATATATATATATATATATATATATATAT 1626
 DB 10353 GTTTAGTTAGTTACCTATATATATATATATATATATATATATATATAT 10412
 QY 1627 ATATTTGATGTAATATCAATATGATATCTGTTTACATATGATATATAT 1686
 DB 10413 AAATTTAGAGAAAGATTTATATATATATATATATATATATATATATAT 10472
 QY 1687 GTTTAATTTTCAATTTGATATATGCGCTTATATATATATATATATATAT 1746
 DB 10473 TTTAAGATCAATTTGCGATTTGTAATTTATATATATATATATATATAT 10532
 QY 1747 CTGTACTGTAAATATATATATATATATATATATATATATATATATAT 1789
 DB 10533 TAAAGGTTTTAGATATATATATATATATATATATATATATATATAT 10575

RESULT 8
 US-10-239-676-164
 Sequence 164, Application US/10239676
 Publication No. US20030082609A1
 GENERAL INFORMATION:
 APPLICANT: OLEK, Alexander
 APPLICANT: PIEPENBROCK, Christian


```

: APPLICANT: BERLIN, Kurt
: TITLE OF INVENTION: diagnosis of Diseases Associated with Gene Regulation
: FILE REFERENCE: 5013.1003
: CURRENT APPLICATION NUMBER: US/10/239,676
: CURRENT FILING DATE: 2002-09-24
: PRIOR APPLICATION NUMBER: PCT/EP01/03968
: DE 10019058.8
: DE 10019173.8
: DE 10032529.7
: DE 10043826.1
: PRIOR FILING DATE: 2001-04-06
: 2000-04-06
: 2000-04-07
: 2000-06-30
: 2000-09-01
: NUMBER OF SEQ ID NOS: 228
: SEQ ID NO 164
: LENGTH: 6030
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (5455)
: US-10-239-676-164

```

Query Match	3.78;	Score 67.6;	DB 9;	Length 6030;
Best Local Similarity	48.48;	Pred. No. 0.0028;		
Matches 222; Conservative	0;	Mismatches 230;	Indels 7;	Gaps 1

OY	1320	ATGCGCCATTATACAGCTAATTTTATATATAGACTTGCCCGCAGTACAATTATATGTAT	1377
Dd	5432	ATATGGTAATAATTTTTTTTATATANAANAATATAAAAATTAGTGGGTATGGTGAGGTTT	5493
OY	1380	ATTTTGCACAAGAGTTCGTGTGACAGCCTATACCATAATTATAACTATAAAACATAGA	1438
Dd	5492	TTCGTAATTTTATGTTATTTGGAGATTGAGTAGAGGAATTTATGCAATTTNAGAGGTAGA	5555
OY	1440	TATAGTGTTTGTACAGTTTAAACCAATGGGAGATCAGAAATATTCATGTATTGACAAAT	1496
Dd	5552	GGTTCGTACGTGAGTCAGCATTTATATATATGTAATTTAGTTAGTTGGCCATGAGAGCAATTTT	5611
OY	1500	GTTTAATATCATCTATATAAT-----CTGCAATTTCTAAGAGGCTATTTTGTCTTT	1555
Dd	5612	GTTTATATAAATAGATATAATATATAATTAATTTAAAAATTTGATTTTTTTAGTTAGTCT	5677
OY	1553	TGCGTCATGACGTATATTTGCAATGGTTGCAATTAATTAATTCATCTGTAATGAATAAT	1612
Dd	5672	TAGTTATATGTTAAGTGTATATATGAGCTGTATGTGTTATAGTATATGTAATGAATAT	5733
OY	1613	TATATGCCAAAATAATATTGATGTTTAAATCAATAGATGATCTCGTTACATGCTCAT	1672
Dd	5732	TGTATGATATAGAAATTTTTTATTTATTTGCAAGAAAGTTTTATTAATAATTTTGGATTTAGAA	5791
OY	1673	ATGCAATATAATCTGTGTATATTTCAATTTTGCATATATGGCCTTAATATTTGATATCTGA	1732
Dd	5792	ATTATATGTAATAATTAATTAATAGTATTAATTAATTTTTTTTAAATATAGATTTAGTTTAA	5851
OY	1733	ATTTTATTTCTCGCTGCTACTGTAATAATATATACCTATA	1771
Dd	5852	ATTATATTTGCTTACCGTATATTTTTTATTAATAATTTTTTA	5890

RESULT 9
US-10-239-676-197
; Sequence 197, Application US/10239676
; Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PREEMERBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

```

? FILE REFERENCE: 5013.1003
? CURRENT APPLICATION NUMBER: US/10/229,676
? CURRENT FILING DATE: 2002-09-24
? PRIOR APPLICATION NUMBER: PCT/EP01/03968
? DE 10019058.8
? DE 10019173.8
? DE 10032529.7
? DE 10043826.1
?
? PRIORITY FILING DATE: 2001-04-06
?
? 2000-04-06
? 2000-04-07
? 2000-06-30
? 2000-09-01
?
? NUMBER OF SEQ ID NOS: 228
?
? SEQ ID NO 197
?
? LENGTH: 7823
?
? TYPE: DNA
?
? ORGANISM: Artificial Sequence
?
? FEATURE:
?
? OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
?
? US-10-239-676-197
?
? Query Match 3.6%; Score 65.4; DB 9; Length 7823;
? Best Local Similarity 45.3%; Pred. NO. 0.0081;
? Matches 237; Conservative 0; Mismatches 286; Indels 0; Gaps 0;

```

[illegible]

QY	1370	TAAATAGTAAATTTTGGACCAAGGAGTCTGGTGGACACACCTACTACTAATTTTAAAGCTATA	1429
Db	2213	GTAATAAATTTGCGGTAATAGCTTTTATTTAAATATATATTAATAAAGAATTGAATGT	2272
QY	1430	AAACAATAGATATAGTGTGGTTCACAGTTTAACATCAATGGAGATCAGAAATTTCTATGTA	1488
Db	2273	AATATGTTATATTAAGAATATAGTTTATATTAAGAATATATTAAGTATATATGTTGAATAGGT	2332
QY	1490	TTGAGAAATGTTTAAATATCATCAATCTATAAATCTGAATTTCTAAGAGCCTATTTTGTTC	1549
Db	2333	TTTAGATTTTATATGTAATAGTTTTTTTACAATATATATTATTGTAGCTTTGATGTGG	2392
QY	1550	TTTTGGCTGATGATGATATTTGAATTGGTTGAAATTAATTAATTAATCTCATTTGTAATAAT	1609
Db	2393	TATTAAGAAGAAATTAATTAATTTGAAAGGCTATATTAATATTTTTTTTTTAAATTATAT	2452
QY	1610	AATATATATGCAAAAATATATATTGATGCTTAAATACAAATAGATGATCTGTTACATGCTT	1669
Db	2453	AAATGCTGAGTTTATATTTTTTTTTTAAATATAGTTAAATTAATTAATTTTGTAGTAGAT	2512
QY	1670	CATATGATATATATATCTGTATTAATTTTCATTTTGATTAATGAGCCTTAATATTTGTGATCT	1729
Db	2513	TAAATGTAAGAAATATGATATGAGATTTAGTTGTTTTTTTAAATTAGATATTTAAACAGA	2572
QY	1730	CTAATTTTATTTCTCTCTGTACTGTAAATAATAATACCTATATATGATATACAAATTTCTT	1789
Db	2573	TTTTTAAATAGTATTTTTTTTATTTAGCTTTTATGTTGGAAGATATATATTTGTTTTT	2632
QY	1790	CACAGATATCTATGCTATTTATTAATTAATTAATTTTACGTG	1832
Db	2633	TAAATAATGTTTAAATATGAGAGTGGTTTATATTTGATATTTTG	2675

```

RESULT 10
US-10-094-240-10
; Sequence 10, Application US/10094240
; Publication No. US20030082637A1
; GENERAL INFORMATION:
; APPLICANT: ZWIEBEL, LAURENCE J.
; TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF
; FILE REFERENCE: N8289
; CURRENT APPLICATION NUMBER: US/10/094,240
; CURRENT FILING DATE: 2001-03-08

```

PRIOR APPLICATION NUMBER: 10/056,405
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: 60/264,649
PRIOR FILING DATE: 2001-01-26
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 4985
TYPE: DNA
ORGANISM: Anopheles gambiae
US-10-094-240-10

Query Match 3.6%; Score 65.2; DB 9; Length 4985;
Best Local Similarity 47.7%; Pred. No. 0.0074;
Matches 225; Conservative 0; Mismatches 243; Indels 4; Gaps 1;

QY 1336 AGTAATTTATATAGACCTTGGCTGCTACATTAATAGTATATTTGGACAGAGCT 1395
DB 2468 AT 2527
QY 1396 TCTGGTGCACAGCTATACCTATATTAAGCTATAAACAATAGATGCTTGTACA 1455
DB 2528 ACTTTAT 2587
QY 1456 GTTAACGATGAGATCAGATATCTATGATGAGAAAATGTTAATCAATCTA 1515
DB 2388 AT 2647
QY 1516 TAAATCTGAATTTCTAAGAGCTATTTGCTTGGCGGAGATGATATTTGAT 1575
DB 2648 AT 2707
QY 1576 TGGTGAAT 1635
DB 2708 GTAT 2767
QY 1636 GTTAACCAAT 1695
DB 2768 TAT 2827
QY 1696 TCATTTGATATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1751
DB 2828 TAT 2887
QY 1752 ACTGTAAT 1803
DB 2888 TTTTAT 2939

RESULT 11
US-10-198-846-8860
Sequence 8860, Application US/10198846
Publication No. US2003009974A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Xu, Yongyao
APPLICANT: Wang, Youzhen
APPLICANT: Steinmann, Kathleen
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
FILE REFERENCE: MRI-049
CURRENT APPLICATION NUMBER: US/10/198,846
CURRENT FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/306,220
PRIOR FILING DATE: 2001-07-18
NUMBER OF SEQ ID NOS: 14084
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8860
LENGTH: 502
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
FEATURE: Homo sapiens

NAME/KEY: misc_feature
LOCATION: 89, 90, 94, 96, 99, 104, 108, 117, 124, 125, 126, 131, 139,
LOCATION: 146, 147, 155, 162, 165, 169, 171, 181, 182, 189, 195, 212,
LOCATION: 214, 219, 226, 228, 237, 239, 241, 242, 244, 246, 249, 252,
LOCATION: 256, 259, 290, 293, 307, 308, 320, 321, 328, 337, 341
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc_feature
LOCATION: 355, 359, 362, 363, 374, 390, 401, 404, 409, 413, 414, 415,
LOCATION: 417, 420, 428, 443, 452, 454, 459, 477, 480, 481, 486, 487,
OTHER INFORMATION: n = A,T,C or G
US-10-198-846-8860

Query Match 3.5%; Score 64; DB 9; Length 502;
Best Local Similarity 42.3%; Pred. No. 0.0048;
Matches 151; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

QY 1473 TCAGAAATATCTATGATATGAGAAAATGTTAATATCAATCTAATCTGAATTTCTA 1532
DB 105 TAAT 164
QY 1533 AGAGGCTATATTTGTTCTTTTGGCTGATGATATATTTGATGCTGATATATA 1592
DB 165 NAATANAATTTAAANNNTTTAATATTTAATCTTTAATATATATATATATAT 224
QY 1593 ATTCATTTGTAATAATATATATATATATATATATATATATATATATATAT 1652
DB 225 TNAATTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 284
QY 1653 ATTCGTTTATGATGATATATATATATATATATATATATATATATATATAT 1712
DB 285 TACTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 344
QY 1713 CTTAATATTTGATCTCAATTTATTTCTCTCTCTCTCTCTCTCTCTCTCT 1772
DB 345 TAAATTTATTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 404
QY 1773 TGTATACATATTTCTTCAGAGAAATCTATGCTATATATATATATATATAT 1829
DB 405 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 461

RESULT 12
US-10-239-676-18
Sequence 18, Application US/10239676
Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239,676
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 18
LENGTH: 5979
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-18

Query Match 3.5%; Score 64; DB 9; Length 5979;
Best Local Similarity 46.6%; Pred. No. 0.013;
Matches 311; Conservative 0; Mismatches 350; Indels 7; Gaps 3;

QY 1077 TAAATGAAGTTGATAGCATTTATTAATTAATTCGACATCATGATGATCTCAGC 1136
DB 1471 TAAATGAAGTTGATAGCATTTATTAATTAATTCGACATCATGATGATCTCAGC 1136
QY 1137 TTGATGATTTTTCGAACTAGATCTATGATGATTTTTCGACAGTTCCTTAATC 1196
DB 1531 TATATATTAATTTGAAATA--TTATATTTTTCGAAAGATTTTTCGATTAAT 1588
QY 1197 ATTATTTCTTACTTCTACCTCTGTTGAACATTTAGAACCTGGATTGGACCC 1256
DB 1589 AATATATTTTATTTATATGATTTAGTAAATTTGAAGTATGATGATTTAAT 1648
QY 1257 AATTTGGAAACGATTCATGATGATGATGATGATGATGATGATGATGATGAT 1312
DB 1649 TATTTTAAATTTAGATTTTAAAGAAATTTTACGTTAAATTTAATGAAATATATAT 1708
QY 1313 TGAAGAGTGGCCATTTATGACATTTTATTAAGACCTTTCCTGACATTA 1372
DB 1709 TAAATATTTTATTTATTTAAATTAAGATATATTTATTTAATTAATGATTA 1768
QY 1373 TAGGATATTTTGGACAGAGATTTCTGACACATTAATTAATTAATTAATTA 1432
DB 1769 TAAATGATATGATGATTTGATTTAAATTTAAATGATTTAAATGATTTAA 1828
QY 1433 CAATGATATGATGATTTGATGATTTTACATGATGATGATGATGATGATGATGAT 1492
DB 1829 TATTAATTTATGATTTTAAAGAAATTTAATTTAATTTTGAATTTAATTTA 1888
QY 1493 AGAAATGTTTA-ATATCATCTATAATCTTGAATTTCTAAGAGCTTATTTGCT 1551
DB 1889 GTATTTTATTTAGATATGATTTATTTAATTTAATTTAATTTAATTTTAT 1948
QY 1552 TTGCTGATGATGATTTTGAATTTGATTTGATTTGATTTGATTTGATTTGAT 1611
DB 1949 TAAATTAATTTGATGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 2008
QY 1612 TTAATGCAAAATTAATTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1671
DB 2009 TTAATTAATTTAATTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 2068
QY 1672 TATGAATTAATTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1731
DB 2069 TATATTTTATTAATTAATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 2128
QY 1732 AATTTTAT 1739
DB 2129 AATTTAAT 2136

RESULT 13
US-10-239-676-195
; Sequence 195, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPERBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; PRIORITY FILING DATE: 2002-09-24
; PRIORITY APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIORITY FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07

2000-06-30
2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 195
; LENGTH: 6620
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-239-676-195

Query Match 3.48%; Score 61.6; DB 9; Length 6620;
Best Local Similarity 44.1%; Pred. No. 0.04;
Matches 308; Conservative 0; Mismatches 384; Indels 6; Gaps 1;

QY 1097 ATTATTTAATTTCTGATCATTCATGATGATGATGATGATGATGATGATGAT 1156
DB 1473 ATATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1532
QY 1157 AAGATCTATGATGATTTTTCGAGATTCATTAATCATTTATTTCTTACTTCT 1216
DB 1533 TATATTTATATATTTATTAATTAATTAATTAATTAATTTTATTTATTAATTAAT 1592
QY 1217 CACCTGTTGAACATTTAGAAACCTGATTTGGAAACCAATTTGGAAACAGATTC 1276
DB 1593 TGGTTTGTAGTGTGTTTATTTAGATTTGTTTAAATTTAGAAATTTGATTA 1652
QY 1277 ATAGTATGAATTAATGAAATTCATATTCGTTTTCGAAAGA-----TGCGCCATT 1330
DB 1653 TAAAGAAATTTAATTAATTAATTAATTAATTAATTTTATTTTATTTGATTTGCTT 1712
QY 1331 ATTACATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1390
DB 1713 GTTTAAATTAATTTGATTAATTAATTAATTAATTAATTAATTTTATTTATTTGCTTAT 1772
QY 1391 GGAATCTGTTGACACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1450
DB 1773 TGTATTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1832
QY 1451 GTACAGTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1510
DB 1833 GTATTAATTTGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGAT 1892
QY 1511 ATCTAATTAATTTGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGAT 1570
DB 1893 TTTATTTATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGAT 1952
QY 1571 TGAATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1630
DB 1953 TTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2012
QY 1631 TTGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1690
DB 2013 TTTAATTTAGTAAAGGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGAT 2072
QY 1691 TAAATTTATTTGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGAT 1750
DB 2073 ATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2132
QY 1751 TACTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1788
DB 2133 TATATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGATTTTTCGAT 2170

RESULT 14
US-10-239-676-51
; Sequence 51, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPERBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

Thu Jun 12 08:30:28 2003

us-09-935-703-2.rnpb

Page 9

OY 1751 TACTGTAATAATAGCTATATGTATACAAATTTCTCAGAGAAAT 1798
| | | | | | | | | | | | | | | | | | | | | |
Db 4290 TTATATATAATAATAATGTTTAAATAATTTTAAAGATATAT 4337

Search completed: June 11, 2003, 13:40:23
Job time : 265 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 09:32:37 ; Search time 114 Seconds
(without alignments)
4928.349 Million cell updates/sec

Title: US-09-935-703-2

Perfect score: 1832
Sequence: 1 ggcacgaggtatgatcatcat.....aaaataaattattctgtg 1832

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/1/lna/5A_COMB.seq: *
2: /cgn2_6/prodata/1/lna/5B_COMB.seq: *
3: /cgn2_6/prodata/1/lna/6A_COMB.seq: *
4: /cgn2_6/prodata/1/lna/6B_COMB.seq: *
5: /cgn2_6/prodata/1/lna/PCUS_COMB.seq: *
6: /cgn2_6/prodata/1/lna/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68.8	3.8	19124	2	US-08-487-826B-13
2	67.2	3.7	837	4	US-08-998-416-288
3	65.2	3.6	20674	4	US-09-641-638-651
4	58.2	3.2	19124	2	US-08-487-826B-13
5	56.8	3.1	636	4	US-08-998-416-1137
6	55	3.0	1634	4	US-08-913-842-20
7	55	3.0	1875	3	US-08-913-842-1
8	54.8	3.0	615	4	US-08-998-416-186
9	54.4	3.0	658	4	US-08-998-416-595
10	52.8	2.9	1511	1	US-07-991-867B-8
11	52.8	2.9	1511	1	US-08-107-755A-8
12	52.8	2.9	1511	2	US-08-544-332-8
13	52.8	2.9	1511	4	US-09-370-861A-8
14	51.8	2.8	5852	1	US-07-867-106-2
15	51.6	2.8	636	4	US-08-998-416-1137
16	51.6	2.8	837	4	US-08-998-416-288
17	51.2	2.8	2110	4	US-09-419-459-1
18	51.2	2.8	5923	4	US-09-064-922-3
19	50.8	2.8	660	1	US-07-991-867B-32
20	50.8	2.8	660	1	US-08-107-755A-32
21	50.8	2.8	660	2	US-08-544-332-32
22	50.8	2.8	660	4	US-09-370-861A-32
23	50.8	2.8	4810	3	US-08-852-629-11
24	50.8	2.8	4838	3	US-08-852-629-15
25	50.6	2.8	688	4	US-08-998-416-972
26	50.6	2.8	1422	1	US-08-319-704-5
27	50.4	2.8	7218	1	US-08-232-463-14

28	50.2	2.7	2430	4	US-08-845-258-3	Sequence 3, Appl1
29	50.2	2.7	2430	4	US-08-845-258-40	Sequence 40, Appl1
30	50.2	2.7	2430	4	US-08-990-571-3	Sequence 3, Appl1
31	50.2	2.7	2430	4	US-08-990-571-40	Sequence 40, Appl1
32	50.2	2.7	2430	4	US-08-723-142A-3	Sequence 3, Appl1
33	50.2	2.7	2430	4	US-08-723-142A-40	Sequence 40, Appl1
34	50.2	2.7	2430	4	US-09-528-784A-3	Sequence 3, Appl1
35	50.2	2.7	2430	4	US-09-528-784A-40	Sequence 40, Appl1
36	50	2.7	711	4	US-08-998-416-786	Sequence 786, App
37	50	2.7	1678	1	US-08-261-677-10	Sequence 10, Appl1
38	50	2.7	1678	1	US-08-384-556A-6	Sequence 6, Appl1
39	50	2.7	1678	2	US-08-331-355A-10	Sequence 10, Appl1
40	50	2.7	1678	5	PCT-US94-12364-10	Sequence 10, Appl1
41	50	2.7	1678	5	PCT-US95-07753-6	Sequence 6, Appl1
42	50	2.7	2327	3	US-09-157-077-1	Sequence 1, Appl1
43	50	2.7	2796	1	US-08-261-677-8	Sequence 8, Appl1
44	50	2.7	2796	1	US-08-384-556A-4	Sequence 4, Appl1
45	50	2.7	2796	2	US-08-331-355A-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-08-487-826B-13/C
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellem, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Knobe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487, 826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29, 655
; REFERENCE/DOCKET NUMBER: NIH121, 001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-0176
; TELEFAX: (619) 235-8550
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13
Query Match 3.88; Score 68.8; DB 2; Length 19124;
Best Local Similarity 48.5%; Pred. No. 1.1e-06;

LOCATION: 1123..3123
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 3124..3297
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 3871..4072
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 5552..5633
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 5758..5880
OTHER INFORMATION: exon 4
NAME/KEY: exon
LOCATION: 5996..6099
OTHER INFORMATION: exon 5
NAME/KEY: exon
LOCATION: 6349..6509
OTHER INFORMATION: exon 6
NAME/KEY: exon
LOCATION: 7379..7522
OTHER INFORMATION: exon 7
NAME/KEY: exon
LOCATION: 8645..8854
OTHER INFORMATION: exon 8
NAME/KEY: exon
LOCATION: 12254..12340
OTHER INFORMATION: exon 9
NAME/KEY: exon
LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc-feature
LOCATION: 17555..20674
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1837
OTHER INFORMATION: 10-510-173 : variable motif ATTTA or TTTT
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623

OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668
OTHER INFORMATION: 10-348-391 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8608
OTHER INFORMATION: 10-349-47 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 8658
OTHER INFORMATION: 10-349-97 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8703
OTHER INFORMATION: 10-349-142 : polymorphic base G or C

NAME/KEY: allele
LOCATION: 8777
OTHER INFORMATION: 10-349-216 : deletion of CTG
NAME/KEY: allele
LOCATION: 8785
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 8926
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492
OTHER INFORMATION: 10-507-321 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 13524
OTHER INFORMATION: 10-507-353 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13535

Query Match 3.6% Score 65.2; DB 4; Length 20674;
Best Local Similarity 49.0%; Pred. No. 7.6e-06;

Matches 236; Conservative 0; Mismatches 238; Indels 8; Gaps 2;

1344 TATTATAGAGCTTGGCTGCTACATTAATGATATTTTGCAAGAGACTGCTGTA 1403
11085 TATTATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTT 11144
1404 CAAGCTATACCTAATTAATGATTAATTAATTAATTAATTAATTAATTAAT 1463
11145 AATATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11204
1464 CAATGAGATCAGAAATTTCTATGATTTGAGAAATGTTAAATCACTAATACTT 1523
11205 AATATTAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11264
1524 GAATTTCTAAGAGCTTATTTGCTTTGGCTGAATGAGTAATTTGAATGCTGAA 1583
11265 TATTAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11324
1584 TATTATTAATTTCTCATGTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1643
11325 TTTATTTAAATTAATTAATTAATTAATTTTAAATTTT---AATATTAAATTTAATTT 11380
1644 AATATAGATGATCTGTTTACATTTGTCATGATTAATTAATTAATTTGATTTTG 1703
11381 AATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTA 11440
1704 AATATGAGCTTAAAT---TTGATGCTAATTTTATTTGCTGCTGCTACTGTA 1759
11441 AATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11500
1760 AATATGAGCTTAAATGATTAATTAATTTCTCAGAAATGATTAATTAATTAATTA 1819
11501 TTTATTTAAATTAATTAATTAATTTTAAATGCTAAGAAATTAATTTTAAATTA 11560
1820 AA 1821
11561 GA 11562

RESULT 4
US-08-487-826B-13
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhuan
APPLICANT: Williams, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPL
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-0176
TELEFAX: (619) 235-8550
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 3.2% Score 58.2; DB 2; Length 19124;
Best Local Similarity 46.4%; Pred. No. 0.0003;
Matches 189; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

1420 AATAGCTTAAACAAATAGATGATGAGTTGTTGACGTTAACTCAATGAGATCAGAA 1479
407 AAAAAAAAAAAAAAAAAAAAAAAAAATTTTATTAATTAATTAATTAATTAATTAATTA 466
1480 ATTCTATGATTAAGAGAAATGTTAATTAATTAATTAATTAATTAATTAATTAATTA 1539
467 GTTCACTTATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 526
1540 TATTTGTTCTTTGGCTCAATGAGTATTTCAATTTGTTGATTAATTAATTAATTTCTCA 1599
527 AATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1659
1600 TTGTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1719
587 AATATTAATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 1766
647 TTTCTTTTATCAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1779
1720 ATTGATCTCAATTTTATTTTCTGCTGTTAATTAATTAATTAATTAATTAATTAATTA 1799
707 CGTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1826
1780 CAATTTCTTCAAGAAATTTCTATGCTAATTAATTAATTAATTAATTAATTTT 1826

QY	1538	CTATTTTGTCTTTGGCGATGAGT -ATATTGAATTGGTGAATTAATTAATTC	1596
Db	470	TAAATTGTTTCTGTAATTTAGCAGAGTAATTTGCTTAATATTTTATTAATAGAA	411
QY	1597	TGATTTGAAAAATTAATTATATGCGCAAAAAATATTGTATGTAATCAAAATAGATGATTC	1656
Db	410	TAAATTTAAGTTTAATTTAATGTCAAAATGTTTATATAATTAATAATTAATTAATTT	351
QY	1657	TGTTTACATGTTCAATGAAATAAATCTGTGTTAAATTCATTGATAATTTGGCCTTT	1716
Db	350	TATATTAGTTTGTATATTAATTAATAGTGCAGTAATTTCAATGATTTCAATATTT	291
QY	1717	AATTTTGTATTCCTAATTTTATTTTCTCTGCTAGCTTAAATTAATAGCTAATATGTA	1776
Db	290	GGAATGAACACTTAAATTTTAATTTTAATTTTAATTTATAGTAGTGCTAATTAATTAAGTAA	231
QY	1777	TAAACAATTTCTCAGAAATTTCTATGCTATTATTTAAATATAATA	1823
Db	230	TAAAAATTTACCTTAATTCATCTTACTAAATTAATTAATAAATAGATA	184

```

RESULT 7
US-08-913-842-1/c
: Sequence 1, Application US/08913842
: Patent No. 6028250
: General Information:
: APPLICANT: OHBA, Toshiharu
: APPLICANT: TAKAHASHI, Shuichi
: APPLICANT: ANMA, Yoshiko
: APPLICANT: ASAZO, Kiyozo
: APPLICANT: KATO, Ikunoskin
: TITLE OF INVENTION: PLANT PROMOTER AND METHOD FOR GENE
: NUMBER OF SEQUENCES: 75
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
: STREET: 419 7th Street N.W., Ste. 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/913,842
: FILING DATE:
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 07-073043
: FILING DATE: 30-MAR-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/JP96/00777
: FILING DATE: 26-MAR-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWDY, Roger L.
: REGISTRATION NUMBER: 25,618
: REFERENCE/DOCKET NUMBER: OHBA-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 628-5197
: TELEFAX: (202) 737-3528
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1875 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
US-08-913-842-1

```

	Query Match	3.08;	Score 55;	DB 3;	Length 1875;
	Best Local Similarity	49.3%;	Pred. No. 0.00082;		
	Matches 171;	Conservative	0;	Mismatches 175;	Indels 1;
QY	1478 ATATTCGATGCTATGCAAAATGTTTAAATTCATCATATTAATCTGATATTTCTAGAGG				1537
DB	530 ATGTCCCTTCTAATTAATCCAAACATCATCCAAACATTAATTAATTTCCAAAGTTTCAAAAAA				471
QY	1538 CTTATTTTGGCTCTTTGGCTGAATGACT-ATATTTGATTTGGTGAATTAATTAATTC				1596
DB	470 TATATGTCTTTCGTATTTAGACAGAGTAATTTGTCTTAATATATTTATTAATAGGA				411
QY	1597 TCATTTGAAAAATTAATTAATATGCCAAAAATATATTTGATGTTAAATCAATAGATGATTC				1656
DB	410 TAAATTTAAAGTTTAATTAATTAATGTCAAAAATGTTTATTAATTAATTAATTAATTTACTATATTT				351
QY	1657 TGATTTACATTTGTCATATGSAATATATATATGCTGTAATTTCAATTTTGATATATGAGCTTT				1716
DB	350 TATATTAATGATTTTGGTATTAATTAATTAATTTGACGTCGATATTTTTCATATGATTTTCATATATT				291
QY	1717 AATATTTGATCTCTAATTTTATTTTCTCTCTGTTACTGTAAATAATAGCTATAAGTA				1776
DB	290 GGATGAGCACTTAAATTAATTAATTAATTAATTAATTAATTAATGAGTCAATTAATTAAGTAA				231
QY	1777 TAAACATTTTCTCAGAGAATTTCTAATGCTATTTTAAATTAATAAATA				1823
DB	230 TAAATATTTACCTCTTAATTCATCTTACTAATTAATTAATTAATTAATTAAGATA				184

RESULT 8
 US-08-998-416-186
 Sequence 186; Application US/08998416
 Patent No. 6239264
 GENERAL INFORMATION:
 APPLICANT: Philippsen, Peter
 APPLICANT: Pohlmann, Rainer
 APPLICANT: Steiner, Sabine
 APPLICANT: Mohr, Christine
 APPLICANT: Wendland, Jürgen
 APPLICANT: Knechtle, Philipp
 APPLICANT: Reibschung, Corinne
 TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYLLII
 TITLE OF INVENTION: AND USES THEREOF
 NUMBER OF SEQUENCES: 1152
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: NO. 6239264artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: No. 6239264th Carolina
 COUNTRY: USA
 ZIP: 27709
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/998,416
 FILING DATE: 24-DEC-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 0016/97
 FILING DATE: 31-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: PF/5-30306/A/GCCL1976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8687
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 186:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 615 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAGI074RP
;
US-08-998-416-186

```

Query Match	3.0%;	Score 54.8;	DB 4;	Length 615;
Best Local Similarity	47.4%;	Pred. No. 0.00065;		
Matches 197; Conservative	0;	Mismatches 217;	Indels 2;	Gaps 1.

QY	1416	AATATTAAGCTATAAACAATAGATATAGTGTTCACAGTTATACATGAGATCA	147
Db	190	AATATATTTAAATGAACTATTTAGTCATGTCCAAATTTTAAATAGTATTA	249
QY	1476	GAAATATCTATGATTTGAGAAAATGTTTAATTCATCTATPAATCTTGAAATTTCTAGA	153
Db	250	TATATTTAGATTAATTAATTTCTTTAAATTAATTTAAATGATTCATCAATTAATTA	309
QY	1536	GGCTATTTTCTGTTTGGCTGAATGATGATATTTGAAATTCGTGATTAATTAAT	159
Db	310	TATATTTATTAATTTGTTTATTAATAAATATATTTTATTAATPAAGATTTAATTAAT	369
QY	1596	CTCATTTGTAATAAATATATATGCCAAATAATATGATGTAATCAATAGATGAT	165
Db	370	TTAAATATGTAATTAATATTTTATTAATAATATCTATTTTAAATTAATTT--ATGTG	427
QY	1656	CTGTTTACATTTGTCATATGATATATAATCTGTGTTAAATTCATTTGATATATGGCTT	171
Db	428	ATTTATATTAATTTAACTTTTATPAAGATTTTATTAATAATTAATTTAACTTAATTC	487
QY	1716	TAAATATTTGATCTCAATTTTATTTTCTCTGTACTGTAAATTAATAGCTAATATGT	177
Db	488	TTATATTAATTAATTTTAATTTATTAATTAATAATATATCTATTTATTTATTTATTTTA	547
QY	1776	ATAACAATTTTCTCAGAGAATCTATGATTTTAAATTAATAATTTTACGT	1831
Db	548	ATTAATTAATTAATTAATTAATTAATTTATATTTATATCAATTTTAAATTAATAATAAT	603

```

RESULT 9
US-08-998-416-595
: Sequence 595, Application US/08998416
: Patent No. 6239264
: GENERAL INFORMATION:
: APPLICANT: Philippsen, Peter
: APPLICANT: Pohlmann, Rainer
: APPLICANT: Steiner, Sabine
: APPLICANT: Mohr, Christine
: APPLICANT: Wendland, Jürgen
: APPLICANT: Knechtle, Philipp
: APPLICANT: Reischung, Corinne
: TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII
: TITLE OF INVENTION: AND US$ THEREOF
: NUMBER OF SEQUENCES: 1152
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: No. 6239264artis Corporation
: STREET: 3054 Cornwallis Road
: CITY: Research Triangle Park
: STATE: No. 6239264th Carolina
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08-998,416
: FILING DATE: 24-DEC-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: CH 0016/97
 FILING DATE: 31-DEC-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 595:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 658 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGIN SOURCE:
 ORGANISM: PAG1408RP

Query Match	3.08;	Score 54.4;	DB 4;	Length 658;
Best Local Similarity	47.48;	Pred. No. 0.00082;		
Matches 163;	Conservative 0;	Mismatches 181;	Indels 0;	Gaps 0

QY	1488	TATGAGAAAATGTTTAAATACAACTTAAATCTTGATATCTCAAGAGGCTTATTCGT	154
Db	186	TATTAACCTTATTCGTGATTAAATATTATTTGCTTATTTATATTTATTTATTTAT	245
QY	1548	TCCTTTGGCGTAAGAGATATTTGAAATTTGGTGAATATTAATTAATCTCAATGTGAAA	1607
Db	246	TATTTTATTAATTTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT	305
QY	1608	ATAATTATATGCCAAAATATATTTGATGATTTAAATCAATAGATGATCTGTTACATG	1667
Db	306	TTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT	365
QY	1668	TTCAATGAATAAATAAATCTGTATTTTCATTTTGTAATTTGGCCTTAATATTTGTAT	1722
Db	366	GTTTGTTATTAATTTTATTTATTTATTTACCTTTTATTTTATGTTATTTTATTTTATTT	425
QY	1728	CTCTAATTTTATTTTCTCTGCTGTAACGTAAATAATATAGCTAATATGTAATCAATTTTC	1787
Db	426	TTTTACTTACTAATATAATATTAATTAATTAATDANCATAGTAATGATATATATATGCTTAA	485
QY	1788	TTGCAAGAAATTCATGCTATTTATTTAAATAATAAATTTACGCT	1831
Db	486	TTCAATTATATAGATGTAATTTTGTGAACATATATATATATATGCTAT	529

RESULT 10
US-07-991-867B-8/C
Sequence 8, Application US/07991867B
Patent No. 5476781
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5476781 Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESSES:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/991.867B

FILING DATE: 12-DEC-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,685
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UFI14.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entomopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (18..218)
NAME/KEY: CDS
LOCATION: complement (234..782)
FEATURE:
NAME/KEY: CDS
LOCATION: 852..1511
US-07-991-867B-8

Query Match 2.9% Score 52.8; DB 1; Length 1511;
Best Local Similarity 49.5%; Pred. No. 0.0025;

Matches 196; Conservative 0; Mismatches 192; Indels 8; Gaps 2;

QY 1416 AATTATAGCTATATAACAATAGATGAGTGTGTACAGTTAACTCAATGAGATCA 1475
|||||
DB 1209 AATTATAGCTATATAACAATAGATGAGTGTGTACAGTTAACTCAATGAGATCA 1150
|||||
QY 1476 GAATATTCCTATGATGAGAAATGTTTAAATCACTATAAATCTTGAATTCCTAGA 1535
|||||
DB 1149 TAATATTTACTTAATATGCTATTTTGTATTAAGATATATCTAAATATGTTTATTTT 1090
|||||
QY 1536 GCGTATTTTGTCTTTGGCT--GAATGAGTATATTTGAATGCTTGAATTAATATA 1593
|||||
DB 1089 ATTTTGTATATAATTTAAATTAATTAATTAATTTGAATTAATTAATTTTAATAT 1030
|||||
QY 1594 TT-----CTCATTGTAATAATAATATATATGCAAAATAATATATTTGATTTAAATCAAT 1647
|||||
DB 1029 TTCTGGAATATATTAATAATATATATATATATATATATATATATATATATATATAT 970
|||||
QY 1648 AGATGATTCGTTTACCTTTTCATATGATTAATATCTGTTAATTTCTTTTGATTA 1707
|||||
DB 969 TAATATATTTTAAAT 910
|||||
QY 1708 TTGGCTTAAATTTTGTATCTCTAATTTTCTCTGTTAGCTAATAATATATAGC 1767
|||||
DB 909 TTAATATTAATTTTAAATATATTTTAAATATATATATATATATATATATATATATAT 850
|||||
QY 1768 TATATATGATATACAAATTTCTTCAAGAAATTTCTAT 1803
|||||
DB 849 TGATATATTTATTTTTCATTTGATTAATTTT 814
|||||

RESULT 11
US-08-107-755A-8/c

Sequence 8, Application US/08107755A
Patent No. 5721352
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Gruidl, Michael E.
TITLE OF INVENTION: No. 5721352e1 Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: U.S.A.
ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/107,755A
FILING DATE: 19-AUG-1993

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,658
FILING DATE: 30-JAN-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991

ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UFI14.C2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entomopoxvirus

FEATURE:
NAME/KEY: CDS
LOCATION: complement (18..218)
NAME/KEY: CDS
LOCATION: complement (234..782)
FEATURE:
NAME/KEY: CDS
LOCATION: 852..1511
US-08-107-755A-8

Query Match 2.9% Score 52.8; DB 1; Length 1511;
Best Local Similarity 49.5%; Pred. No. 0.0025;

Matches 196; Conservative 0; Mismatches 192; Indels 8; Gaps 2;

QY 1416 AATTATAGCTATATAACAATAGATGAGTGTGTACAGTTAACTCAATGAGATCA 1475
|||||
DB 1209 AATTATAGCTATATAACAATAGATGAGTGTGTACAGTTAACTCAATGAGATCA 1150
|||||
QY 1476 GAATATTCCTATGATGAGAAATGTTTAAATCACTATAAATCTTGAATTCCTAGA 1535
|||||
DB 1149 TAATATTTACTTAATATGCTATTTTGTATTAAGATATATCTAAATATGTTTATTTT 1090
|||||
QY 1536 GCGTATTTTGTCTTTGGCT--GAATGAGTATATTTGAATGCTTGAATTAATATA 1593
|||||
DB 1089 ATTTTGTATATAATTTAAATTAATTAATTTTAATTTGAATTAATTAATTTTAATAT 1030
|||||

US-09-370-861A-B

Query Match 2.9%; Score 52.8; DB 4; Length 1511;
Best Local Similarity 49.5%; Pred. No. 0.0025;
Matches 196; Conservative 0; Mismatches 192; Indels 8; Gaps 2;

1416 AATTATAGCTTAAACAAATGATGCTGTTTACAGTTAACTCAATGAGATCA 1475
|||||
1209 AATTATAGCTTAAATATATGATGATTCACAAATTTCTAATGAAATGAGTGA 1150
1476 GAATATCTATGATGAGAAATGTTTATATCAATATAATCTGAAATTTGAGA 1535
|||||
1149 TAATATCTATGATGATGATTTTGTATAGATATCAATATGTTATTTTGA 1090
1536 GCGTATTTTGTCTTTGGCT--GATGAGTATATTTGAATGTTGGAATTAATA 1593
|||||
1089 ATTTGTTATATAAATTTATATTAATTTAAATTTGAAATATTAACCTTTAATAT 1030
1594 TT-----CTCATGTGATAAATATATATGCAAAATATATGATGTAATCAAT 1647
|||||
1029 TTTCGAAATATATTTAAATATATTTATCATATATATATATCAATCTTCTAAT 970
1648 AGAGATCTGTTTACATGTTCAATGATGATATATATCTGTTAATTTCAATTTGATA 1707
1699 TAACATATTTTATATATATATATATATATATATATATATATATATATTTT 910
1708 TTGGCTTAAATTTGATCTCTAATTTATTTCTCTGTTACTGTAATAATATAGC 1767
|||||
909 TTAATTTATATATTTTAAATATATTTATTTAAATATATATATATATATATATAT 850
1768 TATAATGATATACATTTTCTCAGAGATTTCTAT 1803
849 TGATATTTTATTTTCTCATGATTTT 814

RESULT 14

US-07-867-106-2

Sequence 2, Application US/07867106

Patent No. 5389526

GENERAL INFORMATION:

APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz MacKiewicz & No. 5389526rls
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/867.106

PRIOR APPLICATION DATA:

APPLICATION NUMBER: AU PJ 7187

FILING DATE: 02-NOV-1989

ATTORNEY/AGENT INFORMATION:

NAME: Feeney, Joanne Longo

REGISTRATION NUMBER: 35,134

REFERENCE/DOCKET NUMBER: RICE-0002

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-568-3439

TELEFAX: 215-568-3439

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 5852 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ANTI-SENSE: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 2378..5038

NAME/KEY: CDS

LOCATION: 2378..5038

US-07-867-106-2

Query Match 2.8%; Score 51.8; DB 1; Length 5852;
Best Local Similarity 49.4%; Pred. No. 0.0063;
Matches 193; Conservative 0; Mismatches 192; Indels 6; Gaps 2;

1439 ATATGAGTGTGTTGACAGTTTAACCTCAATGAGATCAGAAATATCTATATGAGAAA 1498
|||||
5371 AAATTAAT 5430
1499 TGTTATATCAATCTAATATCTGAAATTTCTAAGAGCTTATTTGCTTTGGCTG 1558
|||||
5431 AATATTTGAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTT 5490
1559 AATGAGTATATTTGAATTTGTTGAATATATATATATATATATATATATATATAT 1618
|||||
5491 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 5550
1619 CAAA-----AATATATTTGATGTTAAATCAATAGATGATTCGTTACATGTTCA 1673
|||||
5551 AAAAACCCTTTACATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 5610
1674 TGATATATATCTGTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 1733
|||||
5611 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAA 5670
1734 TTTTATTTTCTCTGTTACTGTAATAATATATAGCTATATATATATATATATATAT 1793
|||||
5671 TTTAATATATATATTTTAAATATATATATATATATATATATATATATATATATAT 5729
1794 AGAATTTATGCTATTTTAAATATATATAT 1824
5730 TAACAT 5760

RESULT 15

US-08-998-416-1137/C

Sequence 1137, Application US/08998416

Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Philippsen, Peter

APPLICANT: Pohlmann, Rainer

APPLICANT: Steiner, Sabine

APPLICANT: Mohr, Christine

APPLICANT: Wendland, Jurgen

APPLICANT: Knechtle, Philipp

APPLICANT: Redischung, Corinne

TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII

NUMBER OF SEQUENCES: 1152

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6239264artls Corporation

STREET: 3054 Cornwallis Road

CITY: Research Triangle Park

STATE: No. 6239264th Carolina

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8587
INFORMATION FOR SEQ ID NO: 1137:
SEQUENCE CHARACTERISTICS:
LENGTH: 636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1692RP
US-08-998-416-1137

Query Match 2.8%; Score 51.6; DB 4; Length 636;
Best Local Similarity 47.4%; Pred. No. 0.0036;

Matches 219; Conservative 0; Mismatches 239; Indels 4; Gaps 2;

QY 1364 TACAATTAACTGATTTTGGACAGGCTGCTGTCAGACCTATACCTATTATAA 1423
DB 477 TAAATTAATTTTAAATTAATTTCTATATAAAGATTAAATATATATCAACATAATA 418
QY 1424 GCTATTAACATATAGATGTTGTACAGTTTAACCTCAATGAGATCAGAAATTC 1483
DB 417 TTTTAAATAATGATTAATTAATAAATAATTAATTAATTAATTAATTAATTAATC 358
QY 1484 TATGATTAAGAAATGTTAATATCAATCAATCAATCAATCAATCAATCAATCAAT 1543
DB 357 TTTAATAATAATAATATATTTTAAATCAATTAATAATAATAATAATAATAATAATG 298
QY 1544 TTGTTCTTTGGCTGAATGAGTAAAT--TGAATGCTGAATTAATAATAATCTCAAT 1600
DB 297 AATAATCAATTAATTAATTAATAAGAAATAATATATCAATATATTTAATAACTA 238
QY 1601 TGTAAATAATTAATATCCCAAAATATATTGATGTTAATCAATAGATGATTCGTT 1660
DB 237 ATTAAATTTGAACATAGACTAAATAGTATCAATTAATAATTAATTTTAATTAATA 178
QY 1661 TACATTTGTCATGATTAATATATCTGTTAATTTCAATTTGATTAATGGCTTAATA 1720
DB 177 TAAATATTAATATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 118
QY 1721 TTGTATCTCTAATTTATTTTCTCTGCTGTAATGTAATTAATGCTAATGTAATAC 1780
DB 117 ATCAAAATATTAAT--TTTATTAATAATGATTAATAATAGTTTAATAATTAATTAATA 59
QY 1781 AATTTCTTCGAGAGAAATTCATGCTATTAATAATAAAT 1822
DB 58 TATTAATAATAAAGTTTATTAATTAATCTTATTAATAAAT 17

Search completed: June 11, 2003, 11:46:36
Job time : 117 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 07:31:33 ; Search time 430 Seconds
(Without alignments)
9594.560 Million cell updates/sec

Title: US-09-935-703-2

Sequence: 1 ggcacgaggtacatacatcat.....aaaataaataattactgtg 1832

Scoring table: IDENTITY_NIC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N.Geneseq_101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1832	100.0	1832	22	AAH43389
2	1688.6	92.2	3264	23	AAST70229
3	576.4	31.5	1914	24	AAAD24022
4	360.6	19.7	381	22	AAAF1720
5	360.6	19.7	381	22	AAAF1720
6	126.4	6.9	1785	20	AAV81744
7	126.4	6.9	1785	20	AAV81744
8	83.2	4.5	5000	24	AAV81745
9	79.4	4.3	5413	22	AAAF46694

10	78.6	4.3	1692	20	AAV81746
11	54.7	3.9	5447	22	AAAF46694
12	72	3.9	73334	24	AAAF46694
13	72	3.9	73334	24	AAAF46694
14	71.8	3.9	15732	22	AAAF46694
15	71.8	3.9	15732	22	AAAF46694
16	71.4	3.9	15732	24	AAAF46694
17	71.4	3.9	15732	24	AAAF46694
18	71.4	3.9	15732	24	AAAF46694
19	70.8	3.9	15732	24	AAAF46694
20	70.4	3.8	11805	24	AAAF46694
21	70.4	3.8	11805	24	AAAF46694
22	70	3.8	5689	22	AAAF46694
23	70	3.8	5689	22	AAAF46694
24	69.2	3.8	5185	24	AAAF46694
25	68.8	3.8	6365	24	AAAF46694
26	68.8	3.8	6365	24	AAAF46694
27	68.8	3.8	9155	24	AAAF46694
28	68.8	3.8	19124	18	AAAF46694
29	68.8	3.8	19124	21	AAAF46694
30	68.6	3.7	1047	22	AAAF46694
31	67.8	3.7	11047	22	AAAF46694
32	67.8	3.7	11047	24	AAAF46694
33	67.8	3.7	11047	24	AAAF46694
34	67.6	3.7	6030	22	AAAF46694
35	67.6	3.7	6030	24	AAAF46694
36	67.6	3.7	6030	24	AAAF46694
37	67.6	3.7	6292	22	AAAF46694
38	67.6	3.7	11155	24	AAAF46694
39	67.2	3.7	9789	17	AAAF46694
40	67	3.7	9642	24	AAAF46694
41	66.8	3.6	11996	24	AAAF46694
42	66.4	3.6	700	22	AAAF46694
43	66.2	3.6	6175	24	AAAF46694
44	65.8	3.6	6106	22	AAAF46694
45	65.8	3.6	6106	24	AAAF46694

ALIGNMENTS

RESULT 1	AAH43389	standard; CDNA, 1832 BP.
ID	AAH43389	
XX	AAH43389	
AC	15-NOV-2001	(first entry)
DT		
XX		
DE	PTase cDNA clone HATBM23.	
XX		
KW	Protein tyrosine phosphatase; PTase; neural disorder; preservative;	
KW	Alzheimer's disease; Parkinson's disease; immune system disorder;	
KW	rheumatoid arthritis; Grave's disease; muscular disorder; wound healing;	
KW	reproductive disorder; pulmonary disorder; cardiovascular disorder;	
KW	arthritis; infectious disease; viral hepatitis; multiple sclerosis;	
KW	hyperproliferative disorder; neoplasm; epithelial cell proliferation;	
KW	endocrine disorder; diabetes mellitus; transplantation; weight disorder;	
KW	hair loss; skin aging; sunburn; food additive; ss.	
XX		
OS	Homo sapiens.	
XX		
FX		
FT	Key	Location/Qualifiers
FT	CDS	19..222
XX		/*tag= a
XX		/product= "PTase"
PN	W0200166706-A1.	
XX		
PD	13-SEP-2001.	
XX		
PF	22-FEB-2001; 2001WO-US05496.	
XX		

PR 03-MAR-2000; 2000US-0186658.
PR 16-MAR-2000; 2000US-0189881.
XX
PA (HOMA-) HUMAN GENOME SCI INC.

XX Shl Y, Ruben SM;
XX MPI: 2001-565582/63.
DR P-PSDB: AAB47496.

XX Novel isolated protein tyrosine phosphatase polypeptide useful for
XX treating and preventing Alzheimer's disease, rheumatoid arthritis,
XX Grave's disease, arrhythmias, neoplasms, multiple sclerosis and
XX diabetes mellitus

PS Claim 1: Page 276-77; 286pp; English.

XX The sequences given in AAH43389-93 encode protein tyrosine phosphatase
XX (PTPase) polypeptides. The PTPase polypeptides, or the cDNA encoding
XX them, are useful for preventing, treating or ameliorating a medical
XX condition in a mammalian subject. They are useful for diagnosing,
XX preventing or treating neural disorders (e.g. Alzheimer's disease,
XX Parkinson's disease), immune system disorders (e.g. rheumatoid
XX arthritis, Grave's disease), muscular disorder, reproductive
XX disorders, pulmonary disorders, cardiovascular disorders
XX (e.g. arrhythmias), infectious diseases (e.g. viral hepatitis),
XX hyperproliferative disorders (e.g. neoplasms), diseases of the cellular
XX level (e.g. multiple sclerosis), endocrine disorders (e.g. diabetes
XX mellitus), and wound healing and epithelial cell proliferation.
XX They are also useful for preventing hair loss, to prevent skin aging
XX due to sunburn, to maintain organs before transplantation, to modulate
XX mammalian characteristics, to treat weight disorders, to change a
XX mammal's mental or physical state, or as a food additive or
XX preservative.

XX Sequence 1832 BP; 560 A; 299 C; 321 G; 652 T; 0 other;

Query Match 100.0%; Score 1832; DB 22; Length 1832;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCACGAGGTATGATATCATGACGAAATTTAGGCTTTAGAACTTAAGATCTGCTGT 60
DB 1 GGCACGAGGTATGATATCATGACGAAATTTAGGCTTTAGAACTTAAGATCTGCTGT 60
QY 61 GAGTTAACTCTGGGAAATCAACCAAGCAAGAAAGAAAGAAAGATCCGAGATATCTT 120
DB 61 GAGTTAACTCTGGGAAATCAACCAAGCAAGAAAGAAAGATCCGAGATATCTT 120
QY 121 CCATTTCAACATCATGATATGATAGTGGCCCAATGAGAAACAAAGCTTGGCATGTTCA 180
DB 121 CCATTTCAACATCATGATATGATAGTGGCCCAATGAGAAACAAAGCTTGGCATGTTCA 180
QY 181 AAGCAAGAGCACTATCACTTTTGTACGATTTGTCTTGAAAGTCTTGGAAACTCTT 240
DB 181 AAGCAAGAGCACTATCACTTTTGTACGATTTGTCTTGAAAGTCTTGGAAACTCTT 240
QY 241 GACTTGGATTAGAAAGCTTCTGCTCTCACTTGAATTAACAAGGGGTTTGA 300
DB 241 GACTTGGATTAGAAAGCTTCTGCTCTCACTTGAATTAACAAGGGGTTTGA 300
QY 301 CCTCTCAATAAAGACATGTTTGCACGTGCTGAAGGGCTTTGCTATGCATACATCTGC 360
DB 301 CCTCTCAATAAAGACATGTTTGCACGTGCTGAAGGGCTTTGCTATGCATACATCTGC 360
QY 361 TTTCTGTTTATCACTTATTTTCTTTTAAAGCTCCCTGAAGGGCAATATCTTGG 420
DB 361 TTTCTGTTTATCACTTATTTTCTTTTAAAGCTCCCTGAAGGGCAATATCTTGG 420
QY 421 CTTGGGGTATCACTGTTTCTTATGATCTGTAGGCAATATCAAAATTAATCTTCCAC 480
DB 421 CTTGGGGTATCACTGTTTCTTATGATCTGTAGGCAATATCAAAATTAATCTTCCAC 480

QY 481 ATTTCCAGTGAACACAGATGTTTACATTAACGATTGCAAGCTTGGCTATTGTTGAAGG 540
DB 481 ATTTCCAGTGAACACAGATGTTTACATTAACGATTGCAAGCTTGGCTATTGTTGAAGG 540
QY 541 ATTACAGAGCCCATTAAGGATTTAAATATATTCATTAAGATTTATTTGGAAAGTGG 600
DB 541 ATTACAGAGCCCATTAAGGATTTAAATATATTCATTAAGATTTATTTGGAAAGTGG 600
QY 601 CTGAGAGAGCTGAGATTTCCAGAGCTTTGTAGTCTTATTCCTGGAGAAACATAGGC 660
DB 601 CTGAGAGAGCTGAGATTTCCAGAGCTTTGTAGTCTTATTCCTGGAGAAACATAGGC 660
QY 661 CAATATATCATGACCTCTTCCAGAGCTTTTAAAGACAGATGCTATTCATGTTCTTAGT 720
DB 661 CAATATATCATGACCTCTTCCAGAGCTTTTAAAGACAGATGCTATTCATGTTCTTAGT 720
QY 721 AGACCTGCTACTTTTGTGCTGATTTGAATTAACAGGTTTAAAGAGCCGATTAAGGT 780
DB 721 AGACCTGCTACTTTTGTGCTGATTTGAATTAACAGGTTTAAAGAGCCGATTAAGGT 780
QY 781 GGACTAACTTTGACACAAATTTGGCTCATTTCTTCAATTTTCACTGCTGCTCTCT 840
DB 781 GGACTAACTTTGACACAAATTTGGCTCATTTCTTCAATTTTCACTGCTGCTCTCTCT 840
QY 841 ACAGCTGTAGACCAAGACCTGTTGCTGGAGACATTTCAAGATAGGAGAGCTCT 900
DB 841 ACAGCTGTAGACCAAGACCTGTTGCTGGAGACATTTCAAGATAGGAGAGCTCT 900
QY 901 CTGCTGTAACAGTCCAAACATTAATAGATGTTTATTAAGAGCCCAAGAGAGCTTT 960
DB 901 CTGCTGTAACAGTCCAAACATTAATAGATGTTTATTAAGAGCCCAAGAGAGCTTT 960
QY 961 TGCCATGCTGATGTTCTTCTTATCCACCTTACACCTTAACATTAATTAATGTCGCT 1020
DB 961 TGCCATGCTGATGTTCTTCTTATCCACCTTACACCTTAACATTAATTAATGTCGCT 1020
QY 1021 TGTAAAGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1080
DB 1021 TGTAAAGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1080
QY 1081 AAAGTTATATTAAGCATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1140
DB 1081 AAAGTTATATTAAGCATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1140
QY 1141 ATGATTTTCAAACTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
DB 1141 ATGATTTTCAAACTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1201 ATTTCCTTACTTCTCACCTCTGTTGAACATTTTGAAGAGCTGATTTGGAAACCAAT 1260
DB 1201 ATTTCCTTACTTCTCACCTCTGTTGAACATTTTGAAGAGCTGATTTGGAAACCAAT 1260
QY 1261 TTGGAAGACGATTCATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
DB 1261 TTGGAAGACGATTCATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
QY 1321 TGTGGCCATTAATTAAGATTTTATTAAGAGCTTTGCTGCTGATCAATTAATAGTATA 1380
DB 1321 TGTGGCCATTAATTAAGATTTTATTAAGAGCTTTGCTGCTGATCAATTAATAGTATA 1380
QY 1381 TTTTGGACAAGAGGTTCTGCTGACAACTATTAATTAATTAATTAATTAATTAATTAATTAAT 1440
DB 1381 TTTTGGACAAGAGGTTCTGCTGACAACTATTAATTAATTAATTAATTAATTAATTAATTAAT 1440
QY 1441 ATGAGTGTGTTAGAGTTTAACTAAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1500
DB 1441 ATGAGTGTGTTAGAGTTTAACTAAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1500
QY 1501 TTTATATCAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560
DB 1501 TTTATATCAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1560
QY 1561 TGAGTATATTTGAATTTGTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620
DB 1561 TGAGTATATTTGAATTTGTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620

QY 1025 AAAAGCAATATACCTTAACTGCTCTACTCTTTGCCCTTAGCTAACTAATAAG 1084
 |||||||
 Db 2458 AAAAGCAATATACCTTAACTGCTCTACTCTTTGCCCTTAGCTAATAATAAG 2517
 |||||||
 QY 1085 TTGTGATAGGCAATATATATATATATGAGTCATTCATGATGATCTCTATGATG 1144
 |||||||
 Db 2518 TTGTGATAGGCAATATATATATATATGAGTCATTCATGATGATCTCTATGATG 2577
 |||||||
 QY 1145 ATTTTCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1203
 |||||||
 Db 2578 ATTTTCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2637
 |||||||
 QY 1204 TCTTTTACTTCTCACTCTCTGTTGAAACATTTAGAACTGGATTTGGAAACCAATTTTG 1263
 |||||||
 Db 2638 TCTTTTACTTCTCACTCTCTGTTGAAACATTTAGAACTGGATTTGGAAACCAATTTTG 2697
 |||||||
 QY 1284 GAAACCGATTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1323
 |||||||
 Db 2698 GAAACCGATTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2757
 |||||||
 QY 1324 GGCATTTATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1383
 |||||||
 Db 2758 GGCATTTATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2817
 |||||||
 QY 1384 TGGACAAAGAGTCTGCTGACAAAGCTATACCTATATATAGCTATATAACATAGATATG 1443
 |||||||
 Db 2818 TGGACAAAGAGTCTGCTGACAAAGCTATACCTATATATAGCTATATAACATAGATATG 2877
 |||||||
 QY 1444 AGCTTTTGACAGTTAACTCAATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1503
 |||||||
 Db 2878 AGCTTTTGACAGTTAACTCAATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 2937
 |||||||
 QY 1504 AATATCAATCTATTAATCTGATTTCTTAAGAGCTTATTTGCTTTGGCTGATGATG 1563
 |||||||
 Db 2938 AATATCAATCTATTAATCTGATTTCTTAAGAGCTTATTTGCTTTGGCTGATGATG 2997
 |||||||
 QY 1564 GTATATTTGAATGTTGTTGAATATATATATATATATATATATATATATATATATATAT 1623
 |||||||
 Db 2998 GTATATTTGAATGTTGTTGAATATATATATATATATATATATATATATATATATATAT 3057
 |||||||
 QY 1624 AATATATTTGATGTTAAATCAATAGATGATGATGATGATGATGATGATGATGATGATG 1683
 |||||||
 Db 3058 AATATATTTGATGTTAAATCAATAGATGATGATGATGATGATGATGATGATGATGATG 3117
 |||||||
 QY 1684 TCTGTTGATTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1743
 |||||||
 Db 3118 TCTGTTGATTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 3177
 |||||||
 QY 1744 TCTGTTGATTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1803
 |||||||
 Db 3178 TCTGTTGATTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 3237
 |||||||
 QY 1804 GCTATTTATTAATTAATTAATTTACTG 1830
 |||||||
 Db 3238 GCTATTTATTAATTAATTAATTTACTG 3264
 |||||||

RESULT 3

ID AAD24022 standard; cDNA: 1914 BP.
 AC AAD24022:

XX 26-MAR-2002 (first entry)
 DE Human protein phosphatase-4 cDNA.

KW Human; protein phosphatase; PP-4; immune system disorder; AIDS; allergy;
 KW neurological disorder; developmental disorder; Alzheimer's disease;
 KW cell proliferative disorder; Huntington's disease; arteriosclerosis;
 KW renal tubular acidosis; gonadal dysgenesis; cancer; adenocarcinoma;
 KW leukaemia; transgenic animal; gene therapy; ss.

OS Homo sapiens.

FH Key Location/Qualifiers
 FT CDS 203..1465
 FT //tag= a
 FT /product= "Protein phosphatase-4"

W0200196546-A2.

20-DEC-2001.

14-JUN-2001; 2001WO-US19442.

16-JUN-2000; 2000US-212447P.

22-JUN-2000; 2000US-213746P.

29-JUN-2000; 2000US-215210P.

06-JUL-2000; 2000US-216529P.

12-JUL-2000; 2000US-218080P.

21-JUL-2000; 2000US-220117P.

(INCY-) INCYTE GENOMICS INC.

WPI: 2002-090206/12.

P-PSDB: ABE14454.

Novel polypeptide, useful for diagnosing, treating or preventing
 disorders of growth and development, immune system, neurological and
 cell proliferation diseases, comprises cancer protein phosphatase
 polypeptides -

Claim 5; Page 112-113; 116pp; English.

The present sequence is human protein phosphatase (PP)-4 cDNA.
 PP polynucleotide and polypeptide are useful in the diagnosis,
 treatment and prevention of immune system disorders, neurological
 disorders, developmental disorders and cell proliferative disorders.
 Examples of immune system disorders include acquired immune deficiency
 syndrome (AIDS), severe combined immunodeficiency disease (SCID),
 adult respiratory distress syndrome, allergies, amyloidosis,
 anemia, asthma, arteriosclerosis, Crohn's disease, atopic
 dermatitis, diabetes mellitus, emphysema, Goodpasture's syndrome, gout,
 Graves' disease, multiple sclerosis, myasthenia gravis, myocardial or
 pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis,
 psoriasis, Reiter's syndrome, rheumatoid arthritis, Sjogren's syndrome,
 scleroderma, systemic sclerosis, trauma; neurological disorders include
 Alzheimer's disease, Huntington's disease, dementia, epilepsy,
 Parkinson's disease, mental retardation and other developmental
 disorders of central nervous system such as Down's syndrome, cerebral
 palsy, periodic paralysis, mental disorders including mood, anxiety,
 and schizophrenic disorders, seasonal affective disorder such as
 akathisia, amnesia, cataplexy, dyskinesia; developmental disorders
 include e.g. renal tubular acidosis, Duchenne and Becker muscular
 dystrophy, gonadal dysgenesis, hypothyroidism; cell proliferative
 disorders include e.g. actinic keratosis, arteriosclerosis,
 atherosclerosis, bursitis, cirrhosis, hepatitis, psoriasis and
 cancer including adenocarcinoma, leukaemia. The polypeptide and
 polynucleotide are further useful for analyzing proteome of a tissue
 or a cell type, for screening an agonist/antagonist, a compound that
 specifically binds to it or its modulator. The polynucleotide is useful
 for creating knockin humanized animals (pigs) or transgenic animals
 (mice or rats) to model human disease, for generating a transcript image
 of a tissue or cell type, which represents the global pattern of gene
 expression by a particular tissue or cell type.

Sequence 1914 BP; 565 A; 378 C; 438 G; 533 T; 0 other;

Query Match 31.5%; Score 576.4; DB 24; Length 1914;
 Best Local Similarity 99.8%; Pred. No. 5.2e-104;

Matches 577; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 125 TTCAACATCATGATATAGTGGCCAAATGAGAACACAGCTTCGTGATGTTCAACG 184
    |||||||
DB 1337 TTCAACATCATGATATAGTGGCCAAATGAGAACACAGCTTCGTGATGTTCAACG 1396
OY 185 AAGGAGCAGTATCATGTTGTTAGCATATATGCTGGAAGTCTTGGAACCTTGACT 244
    |||||||
DB 1397 AAGGAGCAGTATCATGTTGTTAGCATATATGCTGGAAGTCTTGGAACCTTGACT 1456
OY 245 TTGGATTAAAGAAAGCTTGTGCTCTCACTGGAATTAACCAAGTGGGTTGCACTC 304
    |||||||
DB 1457 TTGGATTAAAGAAAGCTTGTGCTCTCACTGGAATTAACCAAGTGGGTTGCACTC 1516
OY 305 CTCATAAAGAACATGTTGCAATGCTGAGAGGCTTCTGCTATGATACCAATGCTTTC 364
    |||||||
DB 1517 CTCATAAAGAACATGTTGCAATGCTGAGAGGCTTCTGCTATGATACCAATGCTTTC 1576
OY 365 TTGGATTAAAGAAAGCTTGTGCTCTCACTGGAAGTGGGTTGCACTC 424
    |||||||
DB 1577 TTGGATTAAAGAAAGCTTGTGCTCTCACTGGAAGTGGGTTGCACTC 1636
OY 425 GGGTATCAGTCTTACTTATGATCTGCTGAGCAATATCAAAATTAATCTCCACATTT 484
    |||||||
DB 1637 GGGTATCAGTCTTACTTATGATCTGCTGAGCAATATCAAAATTAATCTCCACATTT 1696
OY 485 TCCAGTGAAGACAGATGTTACATAAACGATTCAGCTTGGCTTTGGTTGAAGGATTA 544
    |||||||
DB 1697 TCCAGTGAAGACAGATGTTACATAAACGATTCAGCTTGGCTTTGGTTGAAGGATTA 1756
OY 545 CAGAGCCCAATTAAGATTTAAATATATTCATTAAGATTTTGGAAAGTGGCTGG 604
    |||||||
DB 1757 CAGAGCCCAATTAAGATTTAAATATATATTCATTAAGATTTTGGAAAGTGGCTGG 1816
OY 605 AGAGAGCTGAGATTTCCAGAGCTTTGTAAGTCTTATCTGGGAGACATAAGCCCAT 664
    |||||||
DB 1817 AGAGAGCTGAGATTTCCAGAGCTTTGTAAGTCTTATCTGGGAGACATAAGCCCAT 1876
OY 665 AATCATGACCTCTCCAGGCAATTTTAAAGACAGATGTC 702
    |||||||
DB 1877 AATCATGACCTCTCTCCAGGCAATTTTAAAGACAGATGTC 1914

```

RESULT 4

AAS47150 standard; cDNA; 381 BP.

```

XX AAS47150;
XX
XX AAS47150;
XX
XX 18-DEC-2001 (first entry)
XX
XX Human breast cancer cDNA clone 19464.1.
XX
XX Human; ss; breast cancer protein; tumour; cancer; cytostatic;
XX
XX gene therapy.
XX
XX Homo sapiens.
XX
XX MO200179286-A2.
XX
XX
XX 25-OCT-2001.
XX
XX
XX 12-APR-2001; 2001WO-US12164.
XX
XX
XX 17-APR-2000; 2000US-0551621.
XX
XX 08-JUN-2000; 2000US-0590751.
XX
XX 22-JUN-2000; 2000US-0604287.
XX
XX 20-JUL-2000; 2000US-0620405.
XX
XX
XX (CORI-) CORIYA CORP.
XX
XX
XX Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepler WT;
XX

```

DR WPI; 2001-611721/70.

XX Breast Tumour Proteins and nucleic acids useful for the prevention,
 PT diagnosis and treatment of breast cancer -
 XX
 XX
 XX Claim 1; Page 206; 297pp; English.

XX The invention relates to isolated breast tumour proteins and
 CC nucleic acids that encode them, including immunogenic fragments of the
 CC proteins. Also included are expression vectors expressing the
 CC proteins, transformed cells and antibodies raised against the proteins or
 CC an antigen presenting cell expressing the protein. The proteins and
 CC nucleic acids may be used in the prevention, diagnosis and treatment of
 CC diseases associated with inappropriate breast tumour protein expression,
 CC i.e. breast tumours and breast cancer e.g. by gene therapy. The nucleic
 CC acids and their complements may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acids in
 CC samples, and therefore which patients may be in need of restorative
 CC therapy. The proteins, nucleic acids and antibodies may be used in assays
 CC to identify modulators (e.g. antagonists) of breast tumour protein
 CC expression and activity. The antibodies and antagonists may also be used
 CC to down regulate expression and activity. The antibodies may also be used
 CC as diagnostic agents for detecting the presence of the proteins in
 CC samples (e.g. by enzyme linked immunosorbent assay (ELISA)) and in other
 CC immuno-purification diagnostic techniques. The present sequence is
 CC a cDNA from a breast tumour cDNA library isolated by subtractive
 CC hybridisation against a normal breast cDNA library.

XX Sequence 381 BP; 106 A; 68 C; 51 G; 142 T; 14 other;

XX Query Match 19.7%; Score 360.6; DB 22; Length 381;
 XX Best Local Similarity 95.3%; Pred. No. 9,6e-62;
 XX Matches 363; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```

OY 949 AGAGAGACTTTTCCAGTCTGAGTCTTCTTCCATCCACCTTAACATTAATTA 1008
    |||||||
DB 1 AGAGAGACTTTTCCAGTCTGAGTCTTCTTCCATCCACCTTAACATTAATTA 60
OY 1009 CTTAGTCTGCTTTGTTAAACCAAGTATTAATCTTACCTTGGCTTCTTCCCTT 1068
    |||||||
DB 61 CTTAGTCTGCTTTGTTAAACCAAGTATTAATCTTACCTTGGCTTCTTCCCTT 120
OY 1069 TAGCTAATTAAGTATGATATGATGATATTAATTAATCTGAGCATTCATGATAT 1128
    |||||||
DB 121 TAGCTAATTAAGTATGATATGATGATATTAATTAATCTGAGCATTCATGATAT 180
OY 1129 CTCTCATGTTGATGATATTTTCAAACTAGATCTATGATGTTTTCAGAGTTTC 1188
    |||||||
DB 181 CTCTCATGTTGATGATATTTTCAAACTAGATCTATGATGTTTTCAGAGTTTC 240
OY 1189 ATTAATCATTTATTTCTTACTTCTCACCCTGTTGAAACATTTGAAACCTGGATT 1248
    |||||||
DB 241 ATTAATCATTTATTTCTTACTTCTCACCCTGTTGAAACATTTGAAACCTGGATT 300
OY 1249 GGGAAACCAATTTGGAACCAAGATTCATGATGTAAGTGAAGTGAAGTTCATTTCTG 1308
    |||||||
DB 301 GGGAAACCAATTTGGAACCAAGATTCATGATGTAAGTGAAGTGAAGTTCATTTCTG 360
OY 1309 TTTTGAAGAGATGGCCAT 1329
    |||||||
DB 361 TTTTGAAGAGATGGACCT 381

```

RESULT 5

AAFI7720 standard; cDNA; 381 BP.

```

XX AAF17720;
XX
XX AAF17720;
XX
XX 13-MAR-2001 (first entry)
XX
XX Human breast cancer associated 19464-1 coding sequence.
XX

```

KM Human; breast cancer associated gene; vaccine; diagnosis; therapy; ss.
 XX OS Homo sapiens.
 XX PN WO200060076-A2.
 XX PD 12-OCT-2000.
 XX PF 15-FEB-2000; 2000MO-US05308.
 XX PR 02-APR-1999; 99US-0285480.
 XX PR 23-JUN-1999; 99US-0339338.
 XX PR 02-SEP-1999; 99US-0389681.
 XX PR 03-NOV-1999; 99US-0433826.
 XX PA (CORI-) CORIXA CORP.
 XX PI Yugu J, Dillon DC, Mitcham JL, Xu J, Harlocker SL;
 XX DR WPI; 2001-122627/13.
 XX PT An isolated polypeptide useful for the treatment and diagnosis of
 XX PT tumors e.g. breast cancer comprises at least an immunogenic portion of
 XX PT a breast tumor protein.
 XX PS Claim 6; Page 153; 238pp; English.
 XX CC The present invention provides the coding sequences and some protein
 XX CC sequences of proteins associated with breast cancer in humans. These
 XX CC sequences can be used in the diagnosis and treatment of cancers,
 XX CC particularly breast tumours.
 XX SQ Sequence 381 BP; 106 A; 68 C; 51 G; 142 T; 14 other;
 SQ
 Query Match 19.7%; Score 360.6; DB 22; Length 381;
 Best Local Similarity 95.3%; Pred. No. 9.6e-62;
 Matches 363; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
 QY 949 AGAGGAGCTTGGCAGTCCGATGCTTCCATCCACCCCTAACACTTAATTA 1008
 DB 1 AGAGGAGCTTGGCAGTCCGATGCTTCCATCCACCCCTAACACTTAATTA 60
 QY 1009 CTTAGTGTCTTGTAAAGCAAGTATTAACCTTAACCTTCTTCTTGGCCCTT 1068
 DB 61 CTTAGTGTCTTGTAAAGCAAGTATTAACCTTAACTTGGCTTCTTGGCCCTT 120
 QY 1069 TAGCTAACTAATTAAGTGTGATATGAGCATTTATATATTCGTGAGTCATGAT 1128
 DB 121 TAGCTAACTAATTAAGTGTGATATGAGCATTTATATATTCGTGAGTCATGAT 180
 QY 1129 CTTCTATGTTGATGATTTTCAACCTAAGATCTATGATAGTTTCTTCCAGAGTTC 1188
 DB 181 CTCATCATTTGATGATTTTCAACCTAAGATCTATGATAGTTTCTTCCAGAGTTC 240
 QY 1189 ATTAATCATTTATTTCTTCACTTCTGACCTCTGTTGAACATTTAGAAATGATTT 1248
 DB 241 ATTAATCATTTATTTCTTCACTTCTGACCTCTGTTGAACATTTAGAAATGATTT 300
 QY 1249 GGAACCCCAATTTTGGAAACAGATTCATAGCATGAAATGAACCTTCATTTCTG 1308
 DB 301 GGAACCCCAATTTTGGAAACAGATTCATAGCATGAAATGAACCTTCATTTCTG 360
 QY 1309 TTTTGAAGAAGATGGCCAT 1329
 DB 361 TTTTGAAGAAGATGGACCT 381
 RESULT 6
 AAVB1744
 ID AAVB1744 standard; cDNA; 1785 BP.
 XX AC AAVB1744;
 XX

DT 10-MAR-1999 (first entry)
 XX DE Mouse PTP05 encoding cDNA.
 XX KW PTP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;
 KW type I receptor serine/threonine kinase; cancer; leukemia; lymphoma;
 KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
 KW Parkinson's disease; Huntington's disease; ss.
 XX OS Mus sp.
 XX FH Key Location/Qualifiers
 XX FT CDS 199..1479
 XX FT CDS /tag= a
 XX PD WO9849317-A2.
 XX PF 05-NOV-1998.
 XX PF 27-APR-1998; 98MO-US08439.
 XX PR 23-OCT-1997; 97US-0063595.
 XX PR 28-APR-1997; 97US-0044428.
 XX PR 20-MAY-1997; 97US-0047222.
 XX PR 11-JUN-1997; 97US-0049477.
 XX PR 11-JUN-1997; 97US-0049756.
 XX PR 18-JUN-1997; 97US-0049914.
 XX PA (SUGEN-) SUGEN INC.
 XX PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
 XX PI Markoy D, Omrust S, Peles E, Plozman GD;
 XX DR WPI; 1999-009434/01.
 XX DR P-PSDB; AAW89249.
 XX PT New nucleic acid encoding specific protein tyrosine phosphatases -
 XX PT useful for identifying specific modulators for treatment and
 XX PT prevention of cancer and neurodegenerative disease
 XX PS Claim 2; Page 147; 193pp; English.
 XX CC The present invention describes isolated, enriched or purified nucleic
 XX CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
 XX CC present sequence encodes mouse PTP05. The above proteins, other than
 XX CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
 XX CC substances that modulate their activity (i.e. agonists and antagonists).
 XX CC including NBP) in vivo or in vitro. These substances are used to treat
 XX CC or prevent diseases associated with abnormal signal transduction
 XX CC pathways that involve the proteins, particularly cancer (e.g. Leukemia
 XX CC and Lymphoma), while modulators of ALK-7 (which is a type I receptor
 XX CC serine/threonine kinase) are used to promote neuronal survival,
 XX CC particularly for treating Alzheimer's, Parkinson's or Huntington's
 XX CC diseases. Nucleic acid fragments of the polynucleotides encoding the
 XX CC proteins can be used as probes to identify and clone related sequences;
 XX CC to detect protein-encoded RNA; to generate transgenic animals and in
 XX CC gene therapy (optionally after mutation). Ab are used to determine the
 XX CC proteins.
 XX SQ Sequence 1785 BP; 527 A; 361 C; 408 G; 489 T; 0 other;
 SQ
 Query Match 6.9%; Score 126.4; DB 20; Length 1785;
 Best Local Similarity 64.6%; Pred. No. 1.2e-15;
 Matches 301; Conservative 0; Mismatches 126; Indels 39; Gaps 6;
 QY 102 CAGATACCGAGATATCTTCCATTTCAACATCAGATGATAGTGGCCCAATGAGAGAAC 161
 DB 1328 CTGCCATCGCAGTAAGAACTACTTTTGAACATATGACACGATGAGAGAAC 1387
 QY 162 AACGTTCTGCGATGTTTAAAGAGAGAGATACCTTTGTACGATATTTGCTTG 221
 DB 1388 AGCGCTGTGCGATGTTTAAAGAGAGAGATACCTTTGTATGAAATGTGCTTG 1447

QY 222 AAGTCTTCGGAACCTTCTGACTTTGGATTAGAAAGACTTCTGCTCCTCCTCACTTGA 281
 DB 1448 AAGTCTTCGGAACCTTCTGACTTTGGATTAGAAAGACTTCTGCTCCTCCTCACTTGA 1506
 QY 282 ATTACCAAGTGGTGTGACCTCCTCATTAAGAAACATGTTGACAGTGTGCTGAAGGCTT 341
 DB 1507 GTTACCGAGCAGCTTGGAGCC-----TGAGCCGCTGTGAAGCG-TC 1546
 QY 342 TGCATGCAATACATCTGCTTCTGCTTATTCAGTTTATTTCTTCTTAAAGCTCCCT 401
 DB 1547 TCGGGGCGCTGACAGTCTGCTTC-----TGATTTTCTCTCTGAAGGCTCC 1592
 QY 402 GAAGGCAATATCATTTGCTTGGGCTGATCAGTGTATTTATTTGATCTTCTGAGCA 461
 DB 1593 TGAAGTACACTTACCTGAGGACAGAGTGAAGTCTTCCAC---TTGATCTTCTGAGCA 1649
 QY 462 TATCAAAATTAATCTCCCATTTTCCAGTGAAGACAGATTTACATAAAGCATGACGT 521
 DB 1650 GAGCAAAATTAACCT-CCATGCTCTTACGAAAGGAGTGTGATGAACACCTCCGCT 1708
 QY 522 TGGCTATTTGTTGAAGGATTACAGAGCCCAATTAAGATTAA 567
 DB 1709 TGGCTGTCTGTTTGTGATTACAGAGCTTAATAAAGACTTGA 1754
 RESULT 7
 AAV81745
 ID AAV81745 standard; cDNA; 1896 BP.
 AC AAV81745;
 DT 10-MAR-1999 (first entry)
 DE Mouse PTP05 isoform #1 encoding cDNA.
 KW PTP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;
 KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
 KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
 KW Parkinson's disease; Huntington's disease; ss.
 OS Mus sp.
 FH Key location/Qualifiers
 FT CDS 199..1590
 FT /-tag- a
 XX MO9849317-A2.
 PN 05-NOV-1998.
 PD 27-APR-1998; 98MO-US08439.
 PE 23-OCT-1997; 97US-0063595.
 PR 28-APR-1997; 97US-0044428.
 PR 20-MAY-1997; 97US-0047222.
 PR 11-JUN-1997; 97US-0049477.
 PR 11-JUN-1997; 97US-0049756.
 PR 18-JUN-1997; 97US-0049914.
 XX (SUGEN-) SUGEN INC.
 PA App H, Clary D, Courtenidge SA, Hul TH, Jallal B;
 PI Markby D, Onrust S, Peles E, Plozman GD;
 DR WPI; 1999-009434/01.
 DR P-PSDB; AAW89250.
 XX New nucleic acid encoding specific protein tyrosine phosphatases -
 PT useful for identifying specific modulators for treatment and
 PT prevention of cancer and neurodegenerative disease
 PS Claim 2; Page 147-148; 193pp; English.
 XX

CC The present invention describes isolated, enriched or purified nucleic
 CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
 CC present sequence encodes mouse PTP05. The above proteins, other than
 CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
 CC substances that modulate their activity (i.e. agonists and antagonists,
 CC including NBP) in vivo or in vitro. These substances are used to treat
 CC or prevent diseases associated with abnormal signal transduction
 CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
 CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
 CC serine/threonine kinase) are used to promote neuronal survival,
 CC particularly for treating Alzheimer's, Parkinson's or Huntington's
 CC diseases. Nucleic acid fragments of the polynucleotides encoding the
 CC proteins can be used as probes to identify and clone related sequences;
 CC to detect protein-encoded RNA; to generate transgenic animals and in
 CC gene therapy (optionally after mutation). Ab are used to determine the
 CC proteins.
 SQ Sequence 1896 BP; 572 A; 372 C; 427 G; 525 T; 0 other;
 Query Match 6.9%; Score 126.4; DB 20; Length 1896;
 Best Local Similarity 64.6%; Pred. No. 1.2e-15;
 Matches 301; Conservative 0; Mismatches 126; Indels 39; Gaps 6;
 QY 102 CAGATACCGAGATATCTTCATTTCAATCATCATGATGATAGTGCCCAATGAGAAAC 161
 DB 1439 CTGCCATCGAGAGAAGAACTCTTTTGATGATATGAAACATGAGACCATGAGAAAC 1498
 QY 162 AAGCTTGGCAGTGTTCACAAAGAGAGAGAGATATCATTTGTTCAGATATGCTTGG 221
 DB 1499 AGCCCTGTGGCATGATTCACAAAGAGAGAGAGATGATGATGATGATGATGATG 1558
 QY 222 AAGTCTTCGGAACCTTCTGACTTTGGATTAGAAAGACTTCTGCTCCTCCTCACTTGA 281
 DB 1559 AAGTCTTCGGAACCTTCTGACTTTGGATTAGAAAGACTTCTGCTCCTCCTCACTTGA 1617
 QY 282 ATTACCAAGTGGTGTGACCTCCTCATTAAGAAACATGTTGACAGTGTGCTGAAGGCTT 341
 DB 1618 GTTACCGAGCAGCTTGGAGCC-----TGAGCCGCTGTGAAGCG-TC 1657
 QY 342 TGCATGCAATACATCTGCTTCTGCTTATTCAGTTTATTTCTTCTTAAAGCTCCCT 401
 DB 1658 TCGGGGCGCTGACAGTCTGCTTC-----TGATTTTCTCTCTGAAGGCTCC 1703
 QY 402 GAAGGCAATATCATTTGCTTGGGCTGATCAGTGTATTTGATCTTCTGAGCA 461
 DB 1704 TGAAGTACACTTACCTGAGGACAGAGTGAAGTGTTCAC---TTGATCTTCTGAGCA 1760
 QY 462 TATCAAAATTAATCTCCCATTTTCCAGTGAAGACAGATTTACATAAAGCATGACGT 521
 DB 1761 GAGCAAAATTAACCT-CCATGCTCTTACGAAAGGAGTGTGATGAACACCTCCGCT 1819
 QY 522 TGGCTATTTGTTGAAGGATTACAGAGCCCAATTAAGATTAA 567
 DB 1820 TGGCTGTCTGTTTGTGATTACAGAGCTTAATAAAGACTTGA 1865
 RESULT 8
 ABL55644/c
 ID ABL55644 standard; DNA; 50000 BP.
 AC ABL55644;
 DT 01-JUL-2002 (first entry)
 DE AMEPV genome fragment#2.
 DE AMEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
 KW genetic deficiency disorder; ds.
 OS Amsacta moorei entomopoxvirus.
 PS WO200212526-A2.
 XX

PD 14-FEB-2002.
 PF 10-AUG-2001; 2001WO-US25287.
 XX 10-AUG-2000; 2000US-224479P.
 PR 14-SEP-2000; 2000US-0662254.
 XX (UFL) UNIV FLORIDA.
 PA Moyer RM, Li Y, Bawden AL;
 XX MPI; 2002-227161/28.
 DR Novel recombinant entomopox virus vector useful for delivering
 PT polynucleotide encoding protein to vertebrate cell, comprises
 PT polynucleotide encoding protein operably linked with heterologous
 PT promoter sequence -
 PS Disclosure; Page 150-175; 326pp; English.
 XX The invention relates to a recombinant entomopox virus (EPV) vector,
 CC comprising a polynucleotide encoding a protein operably linked with a
 CC heterologous promoter sequence. The invention also concerns methods for
 CC providing gene therapy for genetic deficiency disorders. Vectors of the
 CC invention are useful for delivering a polynucleotide encoding a protein
 CC to a vertebrate cell preferably a mammalian cell, such as a human cell.
 CC The vector is introduced into the vertebrate cell by infection in a viral
 CC particle, or by transfection, transduction, or injection either in vitro
 CC or in vivo. The vector is useful for the delivery and expression of
 CC biologically useful proteins in gene therapy protocols, and for
 CC delivering large DNA segments for engineering of vertebrate cells.
 CC Polynucleotides of the invention have applications in techniques such as
 CC their use as insertion sites for foreign genes of interest, hybridisation
 CC probes, for chromosome and gene mapping, in PCR technologies, and in the
 CC production of sense or antisense nucleic acids. Vectors of the invention
 CC provide for stable integration and expression of heterologous DNA in host
 CC cells, and are adapted for accepting large heterologous polynucleotide
 CC inserts which can be delivered in an infected or transformed cell and
 CC expressed in a stable fraction. The current sequence represents a
 CC fragment of the genome of the genus B entomopoxvirus from *Amsacta moorei*
 CC (AmePV).
 CC
 XX Sequence 50000 BP; 20514 A; 4505 C; 4614 G; 20367 T; 0 other:
 XX
 XX Query Match 4.58; Score 83.2; DB 24; Length 50000;
 XX Best Local Similarity 47.58; Pred. No. 6.2e-07;
 XX Matches 247; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
 XX
 YY 1310 TTTTGAAGAGTGGCATTATATACAGTAATTTATATAGACTTTCCTGACAACT 1369
 DB 39811 TATAGACATATATTCAGACAAATATTTATTAATTTAAATATGATGATAATTTCA 39752
 YY 1370 TAAATAGTATTTTGGACAGAGGCTGCTGTCAGACAGCTATATTAACCTATA 1429
 DB 39751 ATATCAAAACATATTAATAAAACAGTATTAATGTAATAAAATGTTTGAATAATA 39652
 YY 1430 AAACAATAGATGAGTGTGTTGACAGTTTAACATCAATGAGATCAGATATTTCTATGTA 1489
 DB 39691 AATAGATTAATTAATTAATTAATTTCTTCTATTAATATGATGATAATAATAATAATA 39632
 YY 1490 TTTGAAAGATGTTTATATCAATCAATTAATCTTGAATTTTACAGGCTTATTTGTTTC 1549
 DB 39631 TTTTATTTCTGTCATATTAATAAAGTATCAATTAAGTAATGATAATAAAATCGATTTTTC 39572
 YY 1550 TTTTGGCTGATGATATTTGATTTGCTGATTAATTAATTAATTCATGTTAAAT 1609
 DB 39571 AATCTTCTGATGCTGATTAATAAAATTAATTAATTAATTAATTAATTAATTAATTA 39512
 YY 1610 AATATATGCCAAATATATTTGATGTTAAATCAATAGATGATCTGTTTCAATTTGT 1669
 DB 39511 AAAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 39452
 YY 1670 CATATGAATTAATTAATCTGTGTTAATTTGATTAATTTGATTAATTTGATCT 1729

DB 39451 AATTAATAATAATAATAATAAGTTGATTAATATGATTAATTAATTAATTAATTAATTAAT 39392
 YY 1730 CTAAATTTATTTCTCTCTGTTACTGTAAATAATAGCTAATATGATTAATTAATTTCTT 1789
 DB 39391 CCRATTTATTAATTAATTAATTTATGATGATGATTAATTAATTAATTAATTAATTAAT 39332
 YY 1790 CAGAGAATTCATGCTATTTATTAATAATAATAATTTCT 1829
 DB 39331 ATTATGATTAATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAAT 39292
 RESULT 9
 AAS46694
 ID AAS46694 standard; DNA; 5413 BP.
 XX AAS46694;
 AC
 XX 18-DEC-2001 (first entry)
 DT
 XX Tumour suppressor gene derived chemically modified sequence #417.
 DE
 XX Human: tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX
 PN MO200168912-AZ.
 XX
 PD 20-SEP-2001.
 PF 15-MAR-2001; 2001WO-EP029955.
 XX
 XX 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIC-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX MPI; 2001-602752/68.
 DR
 XX Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer -
 PS Claim 1; SEQ ID NO 417; 27pp; English.
 XX
 XX The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC disulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part

ID AAS46758 standard; DNA: 5447 BP.
 AC AAS46758;
 XX
 DT 18-DEC-2001 (first entry)
 DE Tumour suppressor gene derived chemically modified sequence #482.
 DE
 XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; Cpg dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX Homo sapiens.
 OS
 PN WO200168912-A2.
 XX
 XX 20-SEP-2001.
 PD
 PF 15-MAR-2001; 2001WO-EP02955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 PI Olek A, Plepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer
 XX
 PS Claim 1; SEQ ID No 482; 27pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with Cpg dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes. Sequences with even numbered Seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from Wipo at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 CC Sequence 5447 BP: 1427 A; 183 C; 1297 G; 2540 T; 0 other;
 SQ
 Query Match 4.3%; Score 78; DB 22; Length 5447;
 Best Local Similarity 47.8%; Pred. No. 4.5e-06;
 Matches 260; Conservative 0; Mismatches 280; Indels 4; Gaps 1;
 QY 1285 GAAATGGAATTCATATTCGTTTTCGAAAAAGATGCGCCATTATACAGTAATTTT 1344

Db 2394 GTATTGAAAAATGTAATTTTGAATTCGAATTTTGTATTAATAAAGAAATTTT 2453
 QY 1345 ATTATGACCTTGGCTGCTCAATTAATAGATATTT---TGACAGAGAGTCTCG 1400
 Db 2454 TTAATAAGATGAGAGATTTTAATGATTAATCGTTAGATTGAAGGGAATTTGTA 2513
 QY 1401 TGACAGCTATACCTAATTAATTAAGCTATTAACAAATAGATAGAGTGTTCACGTTA 1460
 Db 2514 AAAAAAAAAAAAAAAAAAAGTTAATTAATGAAGTAAATTAATTTGTAAGTTGT 2573
 QY 1461 ACTCAATGAGATCGCAATATTCATGTATGTGAGAAATGTTAATCACTATAAAT 1520
 Db 2574 TTAATATTGAAATGATTAATATTTGTTTATTAATAAAGTTAAGCTATTGTAAGT 2633
 QY 1521 CTGAAATTTCTAAGAGGCTTATTTGCTTGTGCTGCAATGATTAATTTGAATGCTT 1580
 Db 2634 TTTAATTTTATTAAGTTTAAAAAGTTAATTTTAAAAAGAGGTTAAATTA 2693
 QY 1581 GAATTAATTAATTAATCTCATTTGTAATAATTAATATGCCAAATATATTTGATTA 1640
 Db 2694 TTAGATTTGAATTAATTAATTAATTTATGTTGAATAAATTTTATTTATTTAGATAT 2753
 QY 1641 ATCAATATAGATCTGCTTTACATGTTCAATGATTAATTAATCTGTTAATTTCAAT 1700
 Db 2754 TATTAATTAAGATTTGATTAATTAATTAATTAATTTGATTTATTTTATTAAT 2813
 QY 1701 TTGATTAATTTGCTTAATTAATTTGATTAATTTTATTTTCTGCTGTTACTGTA 1760
 Db 2814 AATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTAATTTTATTTTCTG 2873
 QY 1761 TATAGCTATTAATGATTAATTAATTTTCTGAGAAATTTCAATTTTAAATTA 1820
 Db 2874 TTAATGTAATAATTAATTAATTAATTTTGAATTTTATTTTATTAATAAAGACAAA 2933
 QY 1821 ATAT 1824
 Db 2934 ATAT 2937
 RESULT 12
 ABL92318
 ID ABL92318 standard; DNA: 73334 BP.
 XX
 AC ABL92318;
 XX
 DT 01-JUL-2002 (first entry)
 DE Chemically treated DNA repair gene fragment#64.
 DE
 XX DNA repair; cytosine methylation; PMS2L1; PMS2L2; PMS2L3;
 KW PMS2; L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INP1L1; RFC4;
 KW DITTL; FANCB; XRC8; ataxia telangiectasia; aging; Bloom's syndrome;
 KW Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;
 KW immunodeficiency; trichothiodystrophy; Fanconi's anaemia; solid tumour;
 KW cancer; ds.
 XX
 OS unidentified.
 OS
 PN WO200181622-A2.
 XX
 XX 01-NOV-2001.
 PD
 PF 06-APR-2001; 2001WO-EP03972.
 XX
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Plepenbrock C, Berlin K;

XX WPI; 2002-034446/04.
DR

PT New nucleic acid derived from genes associated with DNA repair, useful
PT for diagnosis, e.g. of ataxia telangiectasia, by determination of
PT cytosine methylation -

PS Claim 1; SEQ ID NO 127; 25pp + sequence listing; English

CC The invention relates to nucleic acids containing a sequence of at least
CC 18 nucleotides of chemically treated DNA of genes associated with DNA
CC repair, and their complements. The invention also relates to nucleic
CC acids comprising at least 18 base pairs of the chemically pretreated DNA
CC of genes associated with DNA repair selected from PMS2L1, PMS2L2,
CC PMS2L7, PMS2L3, PMS2, L4, PMS2L5, PMS2L6, MGMT, MSH, NUDT11, TDG, INPPL1,
CC RRC4, DDTL1, FANCB, or KRCC8. Nucleic acids of the invention and relate
CC oligomers, are useful for diagnosis of diseases associated with gene
CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,
CC Cockayne syndrome, trichothiodystrophy, Fanconi's anaemia, solid tumours,
CC immunodeficiency, trichothiodystrophy, Fanconi's anaemia, solid tumours,
CC and/or by detecting single nucleotide polymorphisms. Determination of
CC individual methylation patterns may allow development of individualised
CC therapies. The sequences given in records AB192132-AB192335 represent
CC chemically pre-treated DNA fragments from genes associated with DNA
CC repair, and their complements.
CC Note: The sequence data for this patent is not represented in the
CC specification, but is based on sequence information supplied by the
CC European Patent Office.

CC Note: The sequence data for this patent is not represented in the
CC specification, but is based on sequence information supplied by the
CC European Patent Office.

SQ Sequence 73334 BP; 18968 A; 858 C; 15329 G; 38179 T; 0 other;

Query Match	3.9%	Score 72:	DB 24;	length 73334;
Best Local Similarity	46.0%;	Pred	0.0001;	
Matches 313;	Conservative	0;	Mismatches 365;	Indels 2;
				Gaps 2

QY	1098	TTATTATATATATCTGACGATCATCAGTGATCTCTCATGTTGTAAGTATTTTCAAACTA	1157
Db	40489	TTATTTTGGTTATGTGTATTTTATAGTGAGGTAGTTTTTGTAGTAGTATTAAG	40558
QY	1158	AGATCATGATAGTTTTTTTCCAGAGTCCATTAATCATTTATTCCTTACTTCTGC	1217
Db	40559	TTGGGTTTGTTTTTTTTTTTAAATTTATTTAGTATTTTGTTTTTAAATGGAT	40618
QY	1218	ACCTCTGTGAAACATTTAGAACATCGATTTGGAAACCAATTTTGGAAAAACAGATTCA	1277
Db	40619	AATTAAATTTATTTATTTAAAGTATTAATGGTAGTAGAATTTATATGTTATTT	40678
QY	1278	TAGTCATGAAATGGAACCTTCATATCTGTTTTTGAAGAAAGTGGCCATTTATACAG	1337
Db	40679	TATTAATTTTTTTTGGTTGTGAGATTTTTTTTTTTTTTTTTTTTTTTTGGAGTTAA	40738
QY	1338	TAAATTTATATGAGACTTGCCCTCGACATTAATAGTATTTTGGACAAGAGTTC	1397
Db	40739	GTCATTTTTTTTNGTGTATATTTTGATTTTTTGTTTTTTATTTTATGTAATTTATAT	40798
QY	1398	TGGTCACAAGCTATACCTAATTTAAACCTATAAACATAGATATGAGCTTTGTACAGT	1457
Db	40799	AGCTTTTTTGTGTGTATTTATNGGATTTATAAAGATGTTTATAGTTATATAAAGG	40858
QY	1458	TTAACTCAATGAGATCGAGATATCTATGTATTTAGAGAAAATGTTAATATCAATCTATA	1517
Db	40859	TT-ATTTTATATGAGTAAATTTTGTATCGAAAAAGAAAATTTTATTTATTTATTTT	40917
QY	1518	AATCTTGAAATTTCTAAGAGCTTATTTGCTCTTTGGCTGAATAGATATTTGAATG	1577
Db	40918	TATTTTGAAGTTT-TGGGTGTTAAATATTAATTTTTTTATATGTAAGTTTTTAATAAA	40978
QY	1578	GTTGAATTAATTAATATTCATCTGATTAATAAATTAATATATGCAAAAATATTTGATGT	1637
Db	40977	TTATCGTAGTATATTTTAATTTGTGTGTTTAAATTTTTTAAABAAAATATAGATGT	41038
QY	1638	TAAATCAATAGATCTGTTTACATGTTTCATATGAAATATAATCTGTATTAATTC	1697

Db 41037 ATATATTATATATAGTATGAGATTTAGAAATTGATGCTATTTGTTTATTAG 41098

QY 1698 ATTTTGATAATTGGCCTTAATTTGATCTCTAATTTATTTTCTCTCTGTACTGTA 1757

Db 41097 TGAGTATATATTTTAGAGCTTTTGCTGTTATGTAATTTTTTTTATTATG 41156

1758 AATAATAGCTATAATGTAT 1777

Db 41157 AAGAAATTTTCTAGTATTT 41176

RESULT 13

ABL34124
ID ABL34124 standard; DNA; 73334 BP.

AC ABL34124;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 2097.

KM Human, immune system disease; cytosine methylation; antisthmatic
 KM antiatherosclerotic; antianemic; cytosatic; nictropic;
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KM antirheumatic; antiarthritic; antididiabetic; antiparasitic;
 KM antinflammatory; cancer; eye disease; arteriosclerotic; anaemia;
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease
 KM gene; ds.

OS Homo sapiens.

PN W0200200928-A2.

03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.
YY

PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.

XX
PA (EPiG-) EPIGENOMICS AG.

XX	Olek A.	Piepenbrock C.
PI		

AA
DR WPI: 2002-130909/17.

AA	PT	Nucleic acid comp
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

PT for diagnosis and treatment of diseases associated with abnormal cytosine methylation -

PS Claim 1; SEQ ID NO 2097; 32pp + Sequence Listing; German.

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/allergic bowel diseases. The present sequence is a gene of the invention.

SQ Sequence 73334 BP; 18968 A; 858 C; 15329 G; 38179 T; 0 other

Query Match	Score 72;	DB 24;	Length 73334;
Best Local Similarity	46.0%;	Pred. No. 0.0001;	
Matches 313; Conservative	0;	Mismatches 365;	Indels 2; Gaps 2;

OY 1098 TTATTATATAATCTGCAGTCATTCATGATATCTTCATCTTTGATGATTTTCAACGTA 1157
 |||||
 Db 40499 TTTATTTAGTATATGTTATTTTATATAGCTGAGCTAAAGTTTTTTGTAGCTAGTATATAG 40555

OY	1158	AGATCTATGATAGTATTTTTCCTCCAGAGTCCATTTAAATCATTTATTTCCCTTCCCTTC	1217
Db	40559	TTGGGTTTGTCTTTTCTTTTCTTTTAAATTTATTTAGTTATTTTGTGTCTTTTAAATGGAT	40618
OY	1218	ACGCTGTTGAAACATTTAGAAACATCGATTTGGGAAACCAATTTTGGAAACCCAGATTTCA	1277
Db	40619	AATTTAATTTTATTTATTTAATTAAGTAATTAATTTGTAAGTAAGATTTATTTATTTGTAATTT	40678
OY	1278	TAGTCATGAAATGGAACCTTCCATATTCGTCTTTTGAAGAAAGATGGCCATTATTACAG	1337
Db	40679	TATTAATTTTCTTTTCTTTTCTTTGTTAGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTAA	40738
OY	1338	TAAATTTTATTAAGACCTTGGCTGCTACATAATTAAGTATATTTTGGACAAGAGACTTC	1397
Db	40739	GTGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTAT	40798
OY	1398	TGCTGACAAAGCTTACCTATTAAGCTATAAAGCAATAGATAGAGTGTGTACACT	1457
Db	40799	AGGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTAAAG	40858
OY	1458	TTAATCTCATGAGATGACAAATATTTCTATGATTTTGAAGAAAGTTTAAATTCATCTATA	1517
Db	40859	TT-ATTTTATATGATAGTAAATTTTATGTCAGAAAGAAATATTTTAAATTTATTTT	40917
OY	1518	AATCTGAAATTTCTAAGAGCTTATTTTGTCTTTGGCTGAATGACTAATTTGAATTC	1577
Db	40918	TATTTTGAATTT-TTGGTGTAAATTAATTAATTTTCTATGCTATGTTTCTTTTAAATTA	40976
OY	1578	GTTGCAATTAATTAATTTCTCATGTTGAAAAATTAATTAATGCCAAAAATTAATTTGATG	1637
Db	40977	TTATCGTACTTTATTTATTTTAAATTTGTTGTTTAAATTTTAAATTAATTAATTAATTAAT	41036
OY	1638	TAAATCAATATGATGATTCCTGTTTACATGTTTCATATGAATTAATTAATTCGTATTAATTC	1697
Db	41037	AATATATATATATATATAGTATTAAGATTTTAAGATTTGATGCTAATTTGTTTATTAATAG	41096
OY	1698	AATTTGATTAATGGCCTTTAAATTTTGTAATCTTAATTTATTTTCTCTGTTACTGTA	1757
Db	41097	TGAGTTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA	41156
OY	1758	AAATATAGCTAATATGAT	1777
Db	41157	AAGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT	41176
RESULT 14			
AAS45388			
XX	AAS45388 standard; DNA: 15732 BP.		
XX	AAS45388;		
DT	18-DEC-2001 (first entry)		
XX	Chemically pretreated genomic DNA associated with cell cycle #47.		
KW	Cell cycle; human; Cpg dinucleotide; cytosine methylation; HIV; aging;		
KW	human immunodeficiency virus; neurodegenerative disorder; solid tumour;		
KW	graft-versus-host disease; glomerular disease; Levy body disease; cancer;		
KW	arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritis;		
KW	immunosuppressive; antitumour; cyostatic; antiarteriosclerotic; ds;		
XX	PCR primer.		
OS	Homo sapiens.		
XX	MO200168911-A2.		
XX	20-SEP-2001.		
XX	15-MAR-2001; 2001WO-EP02945.		
XX	15-MAR-2000; 2000DE-1013847.		
PR	06-APR-2000; 2000DE-1019058.		
PR	07-APR-2000; 2000DE-1019173.		

XX	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	(EPIC-) EPIGENOMICS AG.
XX	Olek A., Piepenbrock C., Berlin K;
XX	WPI; 2001-602751/68.
XX	Designing primers and probes for analysing diseases associated with
PT	cytosine methylation state e.g. arthritis, cancer, aging,
PT	arteriosclerosis comprising fragments of chemically modified genes
PT	associated with cell cycle .
PS	Claim 1; SEQ ID No 93; 28pp; English.
XX	
CC	Sequences AA545296-AA545520 represent chemically pretreated genomic DNA
CC	molecules associated with the cell cycle and specific PCR primers of the
CC	Invention. The sequences are useful for detecting the methylation state
CC	of all CpG dinucleotides in a sequence and therefore for analysing
CC	DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC	of existing diseases or the predisposition to specific diseases can be
CC	ascertained. The parameters may be compared to another set of genetic
CC	and/or epigenetic parameters, the differences serving as basis for
CC	diagnosis and/or prognosis events which are disadvantageous to patients.
CC	The sequences of the invention are useful for the diagnosis and therapy
CC	of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC	aging, glomerular disease, lewy body disease, arthritis,
CC	arteriosclerosis, solid tumours and cancers.
XX	
SQ	Sequence 15732 BP; 4638 A; 70 C; 2672 G; 8352 T; 0 other;
	Query Match 3.9%; Score 71.8; DB 22; Length 15732;
	Best Local Similarity 49.9%; Pred. No. 8.9e-05;
	Matches 208; Conservative 0; Mismatches 207; Indels 2; Gaps 1;
QY	1416 AATTATTAAGCTTAAACAAATGATGATGATGTGGTGACAGTTTAACTCAATGAGATCA 1475
DB	4351 AGTAATTATGAATTTAATTAATTAAGCGTAGAGTAAGTTAGTTATTAATTTAAGAGATA 4410
QY	1476 GAATATTCATGTATTTGAGAAAAGGTTTAATATCAACTATTAATCTGAATTTCTAGA 1533
DB	4411 TTAGATATGAATTTTTTTTAAATTAATTTGAAATGAAGATTAATTAATTTAATGATTTTT 4470
QY	1536 GCCTATATTTGCTCTTTGGCTGGAGTACTATATTTGAATGGTGAATTAATTAATTT 1595
DB	4471 G--TTTTTGTAAAGTTTAGTGATTAAGTTAATTAATTTGGCGAATATGAGATNT 4522
QY	1596 CTCATGTAAAAATATATTTATGCCAAAAATATATTTGATGTTAAATCAATAGATGAT 1655
DB	4529 ATTTAAATATAAATAATTTTATAGTAGATTTATTTTATTAATTAAGATGAGATTAAGTT 4588
QY	1656 CTGCTTACATGTCTCATAGATAAATAACTGGGTAATTTGCATTTGGATTCGCCCTT 1711
DB	4589 TGCTACTAATAGATATTTTATTAACAATTTTGTGTTTGTATTTTGTGCAATAGGAAG 4644
QY	1716 TAATATTTGATCCTAATTTTATTTCTCTCGTACTAGTAATAATATAGCTATPAAGT 1777
DB	4649 GAGGGAATAATTTTGGTGTTTCTTTTTTATTTAATTAATTAAGATTTTATTTAAGCGG 4700
QY	1776 ATAACAATTTCTTCAGAAAGATTCATGCTATTAATTAATAATTAATTTACTGTG 1832
DB	4709 TTATCGGTTTTTAATGATTTATATGTGAAGTTAGTTAATTAAGAAAGTTTGAGATTG 4765
RESULT 15	
ABK28233	
ID	ABK28233 standard; DNA; 15732 BP.
XX	ABK28233;
XX	23-Apr-2002 (first entry)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 11, 2003, 07:34:22 ; Search time 4683 Seconds

(without alignments)
11385.082 Million cell updates/sec

Title: US-09-935-703-2

Perfect score: 1832
Sequence: 1 ggcacgaggtatcatatcat.....aaaataaatactactgtg 1832

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_da:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pla:*
35: em_hlg_rtd:*
36: em_hlg_man:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hlgo_hum:*
40: em_hlgo_mus:*
41: em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1822.4	99.5	2499	9	BC036539	BC036539 Homo sapi
2	1688.6	92.2	2786	9	HSMB00374	AL050040 Homo sapi
3	1645.4	88.8	197389	9	AL358791	AL358791 Human DNA
4	1642.2	89.6	65229	9	AL672108	AL672108 Human DNA
5	576.4	31.5	1914	6	AX354535	AX354535 Sequence
6	360.6	19.7	381	6	AX302888	AX302888 Sequence
7	308	16.8	3090	10	D64141	D64141 Mus musculu
8	170.8	9.3	173324	2	AC130106	AC130106 Rattus no
9	106.6	5.8	55785	2	AC118715	AC118715 Mus muscu
10	97.6	5.3	137889	9	AC073269	AC073269 Homo sapi
11	95	5.2	103737	9	AL603966	AL603966 Human DNA
12	95	5.2	171930	9	AL450334	AL450334 Human DNA
13	86.4	4.7	111861	9	AC069435	AC069435 Homo sapi
14	85.2	4.7	104992	2	AC005504	AC005504 Plasmodu
15	85.2	4.7	169546	2	AC004157	AC004157 Plasmodu
16	84.8	4.6	86827	3	PFMAL3P5	AL034536 Plasmodu
17	84.6	4.6	178785	9	AC068139	AC068139 Homo sapi
18	84.4	4.6	15421	3	PFCEMP1RA	X95275 P. falciparu
19	83.2	4.5	50000	6	AX392734	AX392734 Sequence
20	83.2	4.5	233392	14	AF250284	AF250284 Anascta m
21	81.6	4.5	162445	9	AL158151	AL158151 Human DNA
22	81.6	4.4	178273	2	AC005308	AC005308 Plasmodu
23	80.8	4.4	148075	2	AC110014	AC110014 Homo sapi
24	80.8	4.4	177293	2	AC069525	AC069525 Homo sapi
25	80.6	4.4	17443	3	AC024806	AC024806 Caenorhab
26	80.6	4.4	116696	3	PFMAL3P3	298547 Plasmodu
27	80.4	4.4	72243	9	AL731858	AL731858 Human DNA
28	80.4	4.4	172816	9	AC093899	AC093899 Homo sapi
29	80.2	4.4	127820	9	HSJ612B15	AL049597 Human DNA
30	79.8	4.4	143802	9	AL161431	AL161431 Human DNA
31	79.8	4.4	176174	9	AC007483	AC007483 Homo sapi
32	79.6	4.3	137411	9	AC092066	AC092066 Homo sapi
33	79.4	4.3	5413	6	AX251450	AX251450 Sequence
34	79	4.3	155106	9	AC104069	AC104069 Homo sapi
35	79	4.3	161699	9	AC107300	AC107300 Homo sapi
36	78.8	4.3	185596	9	AC021553	AC021553 Homo sapi
37	78.2	4.3	328866	2	AC107203	AC107203 Homo sapi
38	78	4.3	5447	6	AX251516	AX251516 Sequence
39	77.8	4.2	1192	9	HSJ323759	AJ323759 Homo sapi
40	77.8	4.2	96953	9	AC079621	AC079621 Homo sapi
41	77.6	4.2	6591	8	YSCWTCG06	L36890 Saccharomyc
42	77.6	4.2	15635	3	AB083339	AB083339 Bombyx mo
43	77.6	4.2	15656	3	AB070264	AB070264 Bombyx mo
44	77.6	4.2	152209	9	HS1108011	AL034419 Human DNA
45	77.6	4.2	169422	9	AC093309	AC093309 Homo sapi

ALIGNMENTS

RESULT 1
BC036539
LOCUS BC036539
DEFINITION Homo sapiens, similar to protein-tyrosine-phosphatase homolog DKFZP566K0524.1 - human (fragment), clone MGC:33863 IMAGE:5262458, mRNA, complete cds.
ACCESSION BC036539
VERSION BC036539.1 GI:22328116
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2499)
Strausberg,R.

TITLE Direct Submission
JOURNAL Submitted (16-AUG-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
 DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amdan@systemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

FEATURES
 source
 1. 2499
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:33863 IMAGE:5262458"
 /tissue_type="Brain, hippocampus"
 /clone_id="NIH_MGC_95"
 /lab_host="DH10B"
 /note="Vector: pBluescript"
 192..665
 /codon_start=1
 /product="similar to protein-tyrosine-phosphatase homolog DKPSP56K0524.1 - human (fragment)"
 /protein_id="AAH36539.1"
 /db_xref="GI:22328117"
 /db_xref="LOCUSID:257430"
 /transcript="MSRPDRAPRVNDYEGDSAEEDINFRPSSQENTPSPKVFENVSSEKVLSEINPEHNDYEDVFEPPSSGSDPSMTAGPERRDWSSEDEAA GFSOLSPILSGTRKIVSEGLDQIAQIRPLIFNHEQTAIDCLKILEKNSV"

CDS
 766 a 438 c 486 g 809 t

BASE COUNT
 766 a 438 c 486 g 809 t

ORIGIN
 Query Match 99.5%; Score 1822.4; DB 9; Length 2499;
 Best Local Similarity 99.9%; Pred. No. 4.4e-283;
 Matches 1823; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 960 TAAAGACATGTTGACATGCTGCTGAAGGCTTGGCTATGACATACATCTGCTTTGG 1019
 QY 369 TTATACAGTTATTTCTTTCTTAAAGCTCCGGAAGGCAATATCTTGGCTGGGGT 428
 Db 1020 TTATACAGTTATTTCTTTCTTAAAGCTCCGGAAGGCAATATCTTGGCTGGGGT 1079
 QY 429 GATCAGCTGTTTACTTATGATGCTTGGCAATATCAAAATACCTCCACATTTTCA 488
 Db 1080 GATCAGCTGTTTACTTATGATGCTTGGCAATATCAAAATACCTCCACATTTTCA 1139
 QY 489 GTGAACACATGTTACATATAAGGATTCGAGCTTGGCTATTTGGTGAAGGATTAAGA 548
 Db 1140 GTGAACACATGTTACATATAAGGATTCGAGCTTGGCTATTTGGTGAAGGATTAAGA 1199
 QY 549 GCCCATTAAGATTTAAATATATTCATTAAGATTTATTTGGAAGGTGGCTGAGAG 608
 Db 1200 GCCCATTAAGATTTAAATATATTCATTAAGATTTATTTGGAAGGTGGCTGAGAG 1259
 QY 609 AGCTGAGATTTTCCAGAGCTTTGTAGTCTTATCTGGAGAAACATTAAGCCATATATC 668
 Db 1260 AGCTGAGATTTTCCAGAGCTTTGTAGTCTTATCTGGAGAAACATTAAGCCATATATC 1319
 QY 669 ATGACCTCTTCCAGGATTTTAAAGACATGCTATCTATCTTCTTACTAGAGCTG 728
 Db 1320 ATGACCTCTTCCAGGATTTTAAAGACATGCTATCTATCTTCTTACTAGAGCTG 1379
 QY 729 TACTTTTGTGCTATTTGAATTAACCATTTAAAGAGTCCAGTTAGGCTGACTAAC 788
 Db 1380 TACTTTTGTGCTATTTGAATTAACCATTTAAAGAGTCCAGTTAGGCTGACTAAC 1439
 QY 789 TTTGACACAAATGCTTCCATTTCTTCAATTTTCAATCTGCTGCTTCTTACAGCTGC 848
 Db 1440 TTTGACACAAATGCTTCCATTTCTTCAATTTTCAATCTGCTGCTTCTTACAGCTGC 1499
 QY 849 TTAGACAAAGCTGTGCTGGAGACATTTCAATGATGAGAGACTCCTCCGGTGA 908
 Db 1500 TTAGACAAAGCTGTGCTGGAGACATTTCAATGATGAGAGACTCCTCCGGTGA 1559
 QY 909 ACAATCCAAATTAATATAGATTTTATATAGAAAGCCCAAGAGAGACTTTTCCATGC 968
 Db 1560 ACAATCCAAATTAATATAGATTTTATATAGAAAGCCCAAGAGAGACTTTTCCATGC 1619
 QY 969 CTGAGTCTTCTTCCATCCCACTTACACATTTACTTACTGCTGCTTTTAA 1028
 Db 1620 CTGAGTCTTCTTCCATCCCACTTACACATTTACTTACTGCTGCTTTTAA 1679
 QY 1029 GCAAGTATTCCTTAACCTGCTTACCTTACCTTGGCTTACCTAATAAGTTG 1088
 Db 1680 GCAAGTATTCCTTAACCTGCTTACCTTGGCTTACCTTGGCTTACCTAATAAGTTG 1739
 QY 1089 ATATAGCATTAATATTAATTCAGTCAATTCATGATCTCATGTTGATGATTT 1148
 Db 1740 ATATAGCATTAATATTAATTCAGTCAATTCATGATCTCATGTTGATGATTT 1799
 QY 1149 TTCAAACTAAGATATGATGATTTTCTTCCAGAGTTCATTAATCATTTATTTCTT 1208
 Db 1800 TTCAAACTAAGATATGATGATTTTCTTCCAGAGTTCATTAATCATTTATTTCTT 1859
 QY 1209 TACTTCTCAGCTGCTGTTGAACATTTAGAAAGTGAATTTGGAAACCAATTTGGAAA 1268
 Db 1860 TACTTCTCAGCTGCTGTTGAACATTTAGAAAGTGAATTTGGAAACCAATTTGGAAA 1919
 QY 1269 CCAATTCATAGTATGAAAATGGAACCTTCATATTTCTGTTTGAAGAGATGGCCA 1328
 Db 1920 CCAATTCATAGTATGAAAATGGAACCTTCATATTTCTGTTTGAAGAGATGGCCA 1979
 QY 1329 TTATTCAGTAATTTTATATAGAGCTTGGCTGCTGCAATTAATAGATATTTGGAC 1388
 Db 1980 TTATTCAGTAATTTTATATAGAGCTTGGCTGCTGCAATTAATAGATATTTGGAC 2039
 QY 1389 AAGGAGTTCTGGAGCAAGCTATACCTAATTAAGCTATTAAGATATGATGATGT 1448
 Db 2040 AAGGAGTTCTGGAGCAAGCTATACCTAATTAAGCTATTAAGATATGATGATGT 2099


```

OY 1449 TTGACAGTTTAACTCAATGAGATGACAGATATATCTATGATGTGAGAAAATGTTAATAT 1508
    |||||||
DB 2100 TTGTACAGTTTAACTCAATGAGATGACAGATATATCTATGATGTGAGAAAATGTTAATAT 2159
OY 1509 CAATCAATCAATCTGAAATTTCTAGAGAGCTATATTTGCTTTTGGCTGAATGAGATATA 1568
    |||||||
DB 2160 CAATCAATCAATCTGAAATTTCTAGAGAGCTATATTTGCTTTTGGCTGAATGAGATATA 2219
OY 1569 TTGCAATGCTGGAATATATATATATCTGATGTAATAATATATATATGCCCCAAAATAT 1628
    |||||||
DB 2220 TTGCAATGCTGGAATATATATATATCTGATGTAATAATATATATATGCCCCAAAATAT 2279
OY 1629 ATTGATGTTAATCAATGAGATGATGCTGTTACATGTTGCTATGATGAAATATATATGCT 1688
    |||||||
DB 2280 ATTGATGTTAATCAATGAGATGATGCTGTTACATGTTGCTATGATGAAATATATATGCT 2339
OY 1689 GTTAATTTCAATTTTGAATTTGCTTTATATTTGCTATGCTATGCTATTTTCTCTCT 1748
    |||||||
DB 2340 GTTAATTTCAATTTTGAATTTGCTTTATATTTGCTATGCTATGCTATTTTCTCTCT 2399
OY 1749 GTTACTGTAATAATATAGCTATATATATATATATATTTCTCAGAGAAATTTCTATGCTAT 1808
    |||||||
DB 2400 GTTACTGTAATAATATAGCTATATATATATATATATTTCTCAGAGAAATTTCTATGCTAT 2459
OY 1809 TATTAATAATAATATTTACTGTCG 1832
    |||||||
DB 2460 TATTAATAATAATATTTACTGTCG 2483

RESULT 2
HSM800374 2786 bp mRNA linear PRI 18-FEB-2000
DEFINITION Homo sapiens mRNA; cDNA DKFZp566K0524 (from clone DKFZp566K0524);
ACCESSION AL050040
VERSION AL050040.1 GI:4884281
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2786)
AUTHORS Angorge, W., Winkler, U., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE Direct Subcloning
JOURNAL Submitted (15-MAY-1999) MIPS, Am Klopferstr. 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp566K0524) is available at the R2PD in Berlin.
Please contact the R2PD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@r2pd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES
source
1..2786
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="DKFZp566K0524"
    /tissue_type="kidney"
    /clone_id="566 (synonym: hfkx2). Vector pAMP1; host
    K1-2Blue; sites NotI + SalI"
    /dev_stage="fetal"
    /gene="DKFZp566K0524"
    /cds
    /gene
    /note="strong similarity to protein-tyrosine-phosphatases"
    /codon_start=2
    /product="hypothetical protein"

```

```

/protein_id="CA43248.1"
/db_xref="GI:4884282"
/db_xref="SPTREMBL:O9Y406"
/translation="CPNRCGCGCGCTGAAPQAPQVFNKYNSEKVLIRNPNNDY
EDVEESESSESDPSMTWAKGPFRRDRKNSSEDEBAGSQAISPLISDRKIVSEGL
DOLAQIRPLIPNFEQTAIKDCLKILEKTAAYDIMEFMALLENLNPREFSGNOPS
NREKRNRYDILPYDSTRYPLKSKDYINASYIRIVNCEFEYIATOGPLISIDFW
OMVLENNNSVAMITIREMEGIIKCYHWPISLKRPLEKHFVLENYOILQYFIDR
MFOYVEKSTGSHSYKQLOFTKMPDHGTPASDSFIRKIRARSKSHLNGPVPVHCAG
IGRTGVFLCVHVFCAIVKDCSFINIMDIVAQRREDSGMVOTKEQYHHCYDVLVLEVR
KLTLTD"
polya_signal
polya_site
BASE COUNT 841 a 494 c 550 g 901 t
ORIGIN
Query Match 92.2%; Score 1688.6; DB 9; Length 2786;
Best Local Similarity 99.7%; Pred. No. 1.2e-261;
Matches 1702; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
125 TTCAATCATGATATAGTGGCCCAATGAGAGAAACACTTTCGCATGGTTCAAG 184
    |||||||
DB 1070 TTCAACATCATGATATAGTGGCCCAATGAGAGAAACACTTTCGCATGGTTCAAG 1129
OY 185 AAGAGCAGTATCACTTTGTAAGATATGCTGAAGTCTTCGAAACTTCGACT 244
    |||||||
DB 1130 AAGAGCAGTATCACTTTGTAAGATATGCTGAAGTCTTCGAAACTTCGACT 1189
OY 245 TTGATTAAGAAAGACTCTGCTGCTCACTTGAATTAACAGTGCTTGACCTC 304
    |||||||
DB 1190 TTGATTAAGAAAGACTCTGCTGCTCACTTGAATTAACAGTGCTTGACCTC 1249
OY 305 CTCATTAAGAAAGACTCTGCTGCTCACTTGAATTAACAGTGCTTGACCTC 364
    |||||||
DB 1250 CTCATTAAGAAAGACTCTGCTGCTCACTTGAATTAACAGTGCTTGACCTC 1309
OY 365 TTGATTAAGAAAGACTCTGCTGCTCACTTGAATTAACAGTGCTTGACCTC 424
    |||||||
DB 1310 TTGATTAAGAAAGACTCTGCTGCTCACTTGAATTAACAGTGCTTGACCTC 1369
OY 425 GGGTGATGATGTTTACTTATGATCTGCTGAGCAATATCAATTAATCTCCACATT 484
    |||||||
DB 1370 GGGTGATGATGTTTACTTATGATCTGCTGAGCAATATCAATTAATCTCCACATT 1429
OY 485 TCACATGAACAGATGATTAACAGATGCTGCTGAGCAATATCAATTAATCTCCACATT 544
    |||||||
DB 1430 TCACATGAACAGATGATTAACAGATGCTGCTGAGCAATATCAATTAATCTCCACATT 1489
OY 545 CAGAGCCCAATTAAGATTAATATATATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 604
    |||||||
DB 1490 CAGAGCCCAATTAAGATTAATATATATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1549
OY 605 AGAGAGCTGAGATTTCCAGACTTTGTAAGTCTTATTTGAGAGACATAGGCCAAT 664
    |||||||
DB 1550 AGAGAGCTGAGATTTCCAGACTTTGTAAGTCTTATTTGAGAGACATAGGCCAAT 1609
OY 665 AATCATGACCTTTCAGAGCATTTTAAAGACAGATGCTATTCATCTTCTTAAAGTAGAG 724
    |||||||
DB 1610 AATCATGACCTTTCAGAGCATTTTAAAGACAGATGCTATTCATCTTCTTAAAGTAGAG 1669
OY 725 CCGTACTTTTGTGCTGATTTGAATTAACAGTTTAAAGAGTCACTTAAAGAGTGGAG 784
    |||||||
DB 1670 CCGTACTTTTGTGCTGATTTGAATTAACAGTTTAAAGAGTCACTTAAAGAGTGGAG 1729
OY 785 TAACCTTGAGACAAATTTGCTTCAATTTCTATATTTCAATCTGCTCTCTACAG 844
    |||||||
DB 1730 TAACCTTGAGACAAATTTGCTTCAATTTCTATATTTCAATCTGCTCTCTACAG 1789
OY 845 CTGCTAGACCAAGACTGTTGCTGAGAGCATTTCAATGATTAAGAGAGTCTCTCTG 904
    |||||||
DB 1790 CTGCTAGACCAAGACTGTTGCTGAGAGCATTTCAATGATTAAGAGAGTCTCTCTG 1849
OY 905 GTGAACAGTCCAAACTTAATATGATGTTATATAGAAAGCCAGAGAGACTTTTGGC 964

```


DB 144993 CCAGGAAACAGATGTATACATAAAGCATGCGAGCTGGCTATTTGGTTGAGGAGATAC 144934
 QY 546 AGAGCCCAATTAAGATTTAAATATATCATTAAGATTTATTTGAAAAGGTGGCTGGA 605
 DB 144933 AGAGCCCAATTAAGATTTAAATATATCATTAAGATTTATTTGAAAAGGTGGCTGGA 144874
 QY 606 GAGAGCTGAGATTTCCAGAGCTTTGATAGTTCTATTTCTGGGAGACATTAAGCCATA 665
 DB 144873 GAGAGCTGAGATTTCCAGAGCTTTGATAGTTCTATTTCTGGGAGACATTAAGCCATA 144814
 QY 666 ATCATGACCTTTCCAGAGCTTTTAAAGACAGATGCTATTCATGTTCTTTAGCTAGAGC 725
 DB 144813 ATCATGACCTTTCCAGAGCTTTTAAAGACAGATGCTATTCATGTTCTTTAGCTAGAGC 144754
 QY 726 CTGTACTTTTGTGCTGCTATTTGAATAACCAAGTTTAAAAAGAGCCGATTAGGTGACT 785
 DB 144753 CTGTACTTTTGTGCTGCTATTTGAATAACCAAGTTTAAAAAGAGCCGATTAGGTGACT 144694
 QY 786 AACTTGGACACAAATTTGGCTTCATTTCTACATTTTCATCTGCTGCTCTACAGC 845
 DB 144693 AACTTGGACACAAATTTGGCTTCATTTCTACATTTTCATCTGCTGCTCTACAGC 144634
 QY 846 TGTAGACCAAGACCTGCTGGTGGAGACATTTTCATGATAGAGAGCTCCTCTCGG 905
 DB 144633 TGTAGACCAAGACCTGCTGGTGGAGACATTTTCATGATAGAGAGCTCCTCTCGG 144574
 QY 906 TGAACAGTCCAAATTAATATGATTTTATATGAAGCCCAAGAGAGACCTTTGCCA 965
 DB 144573 TGAACAGTCCAAATTAATATGATTTTATATGAAGCCCAAGAGAGACCTTTGCCA 144514
 QY 966 TGGCTAGTCTTCTGATCCACCCCAACACTAATATTTACTAGTGTGCTTTGTTA 1025
 DB 144513 TGGCTAGTCTTCTGATCCACCCCAACACTAATATTTACTAGTGTGCTTTGTTA 144454
 QY 1026 AAGCAAGATTAACCTTTAACTGCTCTTACTCTTCTGCTTACTAGTAAATTAAGT 1085
 DB 144453 AAGCAAGATTAACCTTTAACTGCTCTTACTCTTCTGCTTACTAGTAAATTAAGT 144394
 QY 1086 TGCATATAGCATTAATATATATTTCTGAGTCATGATGATCTCATCTTTGATGTA 1145
 DB 144393 TGCATATAGCATTAATATATATTTCTGAGTCATGATGATCTCATCTTTGATGTA 144334
 QY 1146 TTTTAACTAAGATTAATATATATTTCTGAGTCATGATGATCTCATCTTTGATGTA 1205
 DB 144333 TTTTAACTAAGATTAATATATATTTCTGAGTCATGATGATCTCATCTTTGATGTA 144274
 QY 1206 CTTTACTTCTCACTCTGTTGAAACATTTAGAACTGGATTTGGAAACCAATTTTGA 1265
 DB 144273 CTTTACTTCTCACTCTGTTGAAACATTTAGAACTGGATTTGGAAACCAATTTTGA 144214
 QY 1266 AACCAGATTCATATGATGATAAATGAAACTTCCATATTTGTTTTGAAAAGATGCG 1325
 DB 144213 AACCAGATTCATATGATGATAAATGAAACTTCCATATTTGTTTTGAAAAGATGCG 144154
 QY 1326 CCATTTATACAGTAATTTATATATGAGACTTGGCTGATACATTAATAGATTTTGG 1385
 DB 144153 CCATTTATACAGTAATTTATATATGAGACTTGGCTGATACATTAATAGATTTTGG 144094
 QY 1386 GACAAGAGTCTGCTGACAACTATACCTAATATATAGCTATAAACAATAGATATGAG 1445
 DB 144093 GACAAGAGTCTGCTGACAACTATACCTAATATATAGCTATAAACAATAGATATGAG 144034
 QY 1446 TGTGTTGACGTTTAACTCAATGAGATCAGAAATTTCTATGATTTGAAAAGCTTTAA 1505
 DB 144033 TGTGTTGACGTTTAACTCAATGAGATCAGAAATTTCTATGATTTGAAAAGCTTTAA 143974
 QY 1506 TATCAATCTAATATCTGAATTTCTAAGAGCTTATTTGCTTTGGCTGATGAGT 1565
 DB 143973 TATCAATCTAATATCTGAATTTCTAAGAGCTTATTTGCTTTGGCTGATGAGT 143914
 QY 1566 ATATTTGAATTTGTTGAATAATTAATTTCTATGTAATAATTAATATGCAAAAA 1625
 DB ATATTTGAATTTGTTGAATAATTAATTTCTATGTAATAATTAATATGCAAAAA 1625

DB 143913 ATATTTGAATTTGTTGAATAATTAATTTCTATGTAATAATTAATATGCAAAAA 143854
 QY 1626 TATATTTGATGTTAAATCAATATAGATATTTCTGTTTACATTTGTCATATGATTAATATC 1685
 DB 143853 TATATTTGATGTTAAATCAATATAGATATTTCTGTTTACATTTGTCATATGATTAATATC 143794
 QY 1686 TGTGTTAATTTCAATTTGATATATTTGGCTTAAATATATTTGATATCTAATTTATTTCTC 1745
 DB 143793 TGTGTTAATTTCAATTTGATATATTTGGCTTAAATATATTTGATATCTAATTTATTTCTC 143734
 QY 1746 TCTGTTACTGTTAAATATATAGTATATATCAATTAACAAATTTCTGAGAGAAFTCATGC 1805
 DB 143733 TCTGTTACTGTTAAATATATAGTATATATCAATTAACAAATTTCTGAGAGAAFTCATGC 143674
 QY 1806 TATATTTAAATTAATATATTTACTGTCG 1832
 DB 143673 TATATTTAAATTAATATATTTACTGTCG 143647

RESULT 4

AL672108/c

LOCUS AL672108 65229 bp DNA linear PRI 20-AUG-2002

DEFINITION Human DNA sequence from clone RP11-142117 on chromosome 10,

complete sequence.

ACCESSION AL672108

VERSION AL672108.7 GI:22449768

KEYWORDS HTG.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 65229)

Clark, S.

Direct Submission

Submitted (20-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquery@sanger.ac.uk

On Aug 22, 2002 this sequence version replaced gi:22415876.

Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >=

30). An attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given

in the feature table with their source databases: EM: EMBL; SW:

SWISSPROT; TR: TrEMBL; WP: WORMPEP; Information on the WORMPEP

database can be found at

http://www.sanger.ac.uk/projects/C-elegans/wormpep

This sequence

was generated from part of bacterial clone contigs of human

chromosome 10, constructed by the Sanger Centre Chromosome 10

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr10

RP11-142117 is from the library RPc1-11.1 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBAC3.6.

FEATURES

source

1..65229

/organism="Homo sapiens"

BASE COUNT 565 a 378 c 438 g 533 t
 ORIGIN
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 7476861C1"

Query Match 31.5%: Score 576.4; DB 6; Length 1914;
 Best Local Similarity 99.8%; Pred. No. 3.4e-83;
 Matches 577; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

125 TTCACATCATGATATAGTGGCCCAATGAGAGACACGTTCTGCGATGTTCAACG 184
 1337 TTCACATCATGATATAGTGGCCCAATGAGAGACACGTTCTGCGATGTTCAACG 1396
 185 AAGGACGATATACCTTTTTCAGATATGCTTGAAGTCTTGGAACTCTGACT 244
 1397 AAGGACGATATACCTTTTTCAGATATGCTTGAAGTCTTGGAACTCTGACT 1456
 245 TTGGATTAGAAAGACTTCTGCTCTCCTGCTGAAATATACCAAGTGGTTGGACCTC 304
 1457 TTGGATTAGAAAGACTTCTGCTCTCCTGCTGAAATATACCAAGTGGTTGGACCTC 1516
 305 CTCATTAAGAACATGTTTGCACCTGCTGAAAGGCTTGGCTATGCAATCTGCTTC 364
 1517 CTCATTAAGAACATGTTTGCACCTGCTGAAAGGCTTGGCTATGCAATCTGCTTC 1576
 365 TTGGATTAGATTTTCTTCTTAAAGCTCCCTGAAAGGCAATATCATTTGGCTTG 424
 1577 TTGGATTAGATTTTCTTCTTAAAGCTCCCTGAAAGGCAATATCATTTGGCTTG 1636
 425 GGGTGATCACTGTTTACTTATGATCTGCTAGGCAATATCAAAATACTCCACATTT 484
 1637 GGGTGATCACTGTTTACTTATGATCTGCTAGGCAATATCAAAATACTCCACATTT 1696
 485 TTCAGTAAACAGATGTTTACTTAAAGCATGTCAGCTTGGCTATTTGGTGAAGGATTA 544
 1697 TTCAGTAAACAGATGTTTACTTAAAGCATGTCAGCTTGGCTATTTGGTGAAGGATTA 1756
 545 CAGAGCCCAATTAAGATTTAAATATTCATTAAGATTTTAAAGTGGCTGG 604
 1757 CAGAGCCCAATTAAGATTTAAATATTCATTAAGATTTTAAAGTGGCTGG 1816
 605 AGAGAGCTGAGGATTTCCAGGACTTGTAGTCTTATTTGGGAGAACATTAAGCCCAT 664
 1817 AGAGAGCTGAGGATTTCCAGGACTTGTAGTCTTATTTGGGAGAACATTAAGCCCAT 1876
 665 AATCATGACCTCTCCAGGACTTTTAAAGACAGATGTC 702
 1877 AATCATGACCTCTCCAGGACTTTTAAAGACAGATGTC 1914

RESULT 6
 AX302888 381 bp DNA linear PAT 30-NOV-2001
 LOCUS AX302888
 DEFINITION Sequence 208 from Patent WO01/9286.
 ACCESSION AX302888
 VERSION AX302888.1 GI:17383386
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1
 JIANG, Y., DILLON, D.C., MITCHAM, J.L., XU, J., HARLOCKER, S.L. and
 HEPLER, W.T.
 TITLE Compositions and methods for the therapy and diagnosis of breast
 cancer
 JOURNAL Patent: WO 01/9286-A 208 25-OCT-2001;
 CORIXA CORPORATION (US)
 FEATURES
 source 1..381
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

BASE COUNT 106 a 68 c 51 g 142 t 14 others
 ORIGIN

Query Match 19.7%: Score 360.6; DB 6; Length 381;
 Best Local Similarity 95.3%; Pred. No. 2.2e-48;
 Matches 363; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

949 AGAGAGACTTTGGCAGCTGAGTCTTCTGATCCACCTTAACATTAATATTA 1008
 1 AGAGAGAGATTTGGCAGCTGAGTCTTCTGATCCACCTTAACATTAATATTA 60
 1009 CTGAGTCTGCTTTGTTAAAGCAATATTAATTAATTAATTAATTAATTAATTA 1068
 61 CTGAGTCTGCTTTGTTAAAGCAATATTAATTAATTAATTAATTAATTAATTA 120
 1069 TGGTAACTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1128
 121 TGGTAACTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 180
 1129 CTCATATGTTGATGATTTTTCACCAATGATGATGATGATGATGATGATGATG 1188
 181 CTCATATGTTGATGATTTTTCACCAATGATGATGATGATGATGATGATGATG 240
 1189 ATTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1248
 241 ATTAATATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300
 1249 GGAACCCCAATTTTGGAAACAGATTCATGATGATGATGATGATGATGATGAT 1308
 301 GGAACCCCAATTTTGGAAACAGATTCATGATGATGATGATGATGATGATGAT 360
 1309 TTTTGGAAAGATGAGCCAT 1329
 361 TTTTGGAAAGATGAGCCAT 381

RESULT 7
 D64141 3090 bp mRNA linear ROD 13-FEB-1999
 LOCUS D64141
 DEFINITION Mus musculus mRNA for protein-tyrosine-phosphatase, complete cds.
 ACCESSION D64141
 VERSION D64141.1 GI:265457
 KEYWORDS protein-tyrosine-phosphatase.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (sites)
 Ohsugi, M., Kuramochi, S., Matsuda, S. and Yamamoto, T.
 TITLE Molecular cloning and characterization of a novel cytoplasmic
 protein-tyrosine phosphatase that is specifically expressed in
 spermatocytes

JOURNAL J. Biol. Chem. 272 (52), 33092-33099 (1997)
 MEDLINE 98070510
 REFERENCE 2 (bases 1 to 3090)
 AUTHORS Ohsugi, M.
 TITLE Direct Submission
 JOURNAL Submitted (15-SEP-1995) Miho Ohsugi, The University of Tokyo,
 Department of Oncology; 4-6-1 Shirokanedai, Minato-ku, Tokyo 108,
 Japan (E-mail:mohsugi@leims.u-tokyo.ac.jp, Tel.03-5449-5305,
 Fax:03-5449-5413)

FEATURES
 source 1..3090
 Location/Qualifiers

1..3090
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /tissue-type="testis"
 218..1498
 /cfeature="3.1.3.48"
 /codon_start=1
 /product="protein-tyrosine-phosphatase"
 /protein_id="BAA23761.1"
 /db_xref="GI:265458"
 /translation="MSSPRVVRGRTGRDNDDEBNSGNLNLRLNSLPSSQKMPTRPI"

ENKNSSENVKPSHHSFSDKVELYYPELESDETDENVDSRLBNKNSSETA
 GSKTSPVLSSSRLSKDTEVSSEKELQMLQPLRLNNSRSRMRCLNTLOAK
 BELDIREFLEQMTLPDDFNSGNTLQNDKNRDLIDYSTRVLGNKDIANS
 YRIVNHEEYVIAFOGPELIEDFQVLENNKNCVIMVIREIDCGYIKCSYMP
 ISLKEPLEEHEEYVLEFHVHVOYFVRYVQIYKSKSGCVKHLQDPDGTSA
 SADFPIKRYRYRKRSHITPLVHCSAGVRTGVICVDVVFSAIEKNYSPDINIVT
 QNRKORCMIOREQOFCEIYEVNQLALX*

BASE COUNT 904 a 603 c 647 g 936 t
 ORIGIN

Query Match 16.8%; Score 308; DB 10; Length 3090;
 Best Local Similarity 59.3%; Pred. No. 3.3e-40;
 Matches 1079; Conservative 0; Mismatches 575; Indels 166; Gaps 26;

102 CAGATACGAGATATCTTCATTCATCATGATATAGTGGCCAAATGAGAGAC 161
 1347 CTGGCAGGAGAGAACTACTCTTTGACATATGACATAGTACCCAGATGAGAAAGC 1406
 162 AACGTTCTGGCATGTTCAACAGAGAGAGATGACATTTTGTACATATGTCCTG 221
 1407 AGGCTGTGGCATGATTCACAAAGAGAGAGATGACATTTTGTATGAAATGTGCTG 1466
 222 AAGTTCTGGCACTCTGACTTTGATTAAGAAAGATCTGTGCTGCTCACTTGA 281
 1467 AAGTTCTGCAAACTCTGCTTTGTATTAAGAGAGATCTG-CGCGTGTCCCTGAG 1525
 282 ATTACCAAGTGGTGTGACCTCTCTCATTAAGAAATGTTGACATGCTGTGAGAGGCTT 341
 1526 GTTACCGAGAGAGCTTGAGAGC-----TGAGCCGCTGTGAGAGCG--TC 1565
 342 TGTATGACATACATCTGCTTTCTTGTATTAATTTCTTCTTAAGAGCTCT 401
 1566 TGGGGGCGGACATCTGCTTC-----TGATTTCTCTGTGAGAGCTCT 1612
 402 GAAGGCAATATCATTTGGCTTGGGCTGATGATGTTACTTATGATCTTGTAGGCA 461
 1613 GAAGGTGACACTACTGGGACAGAGTGAAGTGAAGCTGTTCCAG--TGATCTTTCTGAACA 1669
 462 TATCAAAATATCTCCACATTTTCCAGTGAACAGATGTTACATTAAGAGATGACAGT 521
 1670 GACCAAAATACCT--CCATGCTCTTACGGAAGGAGTGAAGAAACCAACCTCCGCT 1728
 522 TGGCTATTTGGTGAAGGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 581
 1729 TGGCTATTTGGTGAAGGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1787
 582 ATTTATTTGGAAGTGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 641
 1788 ATTTATTTGGAAGTGGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1846
 642 TTCTGGAGAGAGATGAGCCAAATA--TCATGACCTTCCAGGCAATTTTAAGACAGA 698
 1847 CTCTGGAGAGAGATGAGCCAAATA--TCATGACCTTCCAGGCAATTTTAAGACAGA 1906
 699 T--GTCTATTCATTTCTTTAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 756
 1907 TTGGAGAGAGAGATGAGTGTGTAGAGATTAATTAATTAATTAATTAATTAATTAAT 1963
 757 GTTTAAAGAGAGAGATGAGTGTGTAGAGATTAATTAATTAATTAATTAATTAATTAAT 816
 1964 --TCAAAGAGAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2021
 817 ACATTTTCAATA-----CTGCTGCTTCTCTAC 842
 2022 ATACTTCATGCTGCCAG 2081
 843 AGCTGCTAGACCAAG 901
 2082 AGCAGTACAG 2141
 902 -----TCGGTGAACAGTCAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 942
 2142 TTACTATCCCTGAACAG 2201

943 AGCCAG 998
 2202 ATTC--AGTGAAG 2259
 999 TAACATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1058
 2260 TAACATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2319
 1059 CTTGCCCCCTTAAGTAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1118
 2320 CTTTATGAG 2373
 1119 TTCAATGATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1175
 2374 TTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2433
 1176 TT-----TCAG 1230
 2434 TTTTCTCTAG 2479
 1231 CATTAAG 1290
 2480 TATTAAG 2538
 1291 GGAAGCTTCCATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1350
 2539 AAAAATTTCCAG 2596
 1351 GGACTTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1410
 2597 -----GAGCTGAG 2649
 1411 TACCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1464
 2650 CTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2709
 1465 AATGAG 1524
 2710 TGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2769
 1525 AATTTCTAAG 1584
 2770 TGGTAAAG 2829
 1585 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1643
 2830 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2889
 1644 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1692
 2890 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2949
 1693 AATTTCTAAG 1752
 2950 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3009
 1753 CTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1812
 3010 CTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3069
 1813 AAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1872
 3070 AAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3089

RESULT 8
 AC130106 173224 bp DNA linear HTG 08-AUG-2002
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-17N23, *** SEQUENCING IN PROGRESS
 AC130106
 AC130106.1 GI:22138328

KEYWORDS

SOURCE

HTG: HTGS, PHASE1.
Norway rat.

ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE

1 (bases 1 to 173224)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Albrooks, S.L., Amaral, H.C., Are, J.R., Ayele, M., Banks, T., Barbata, J., Benton, J., Bimarge, K., Blankenburg, K., Bonin, D., Bouck, J., Boyle, S., Blevins, K., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.E., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthett, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Hawlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homsl, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratoch, J., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louseged, H., Lora, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., McWhinley, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokken, S., Ogih, M., Okunoda, G., Oragunde, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pul, L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojudo, I., Rolfe, M., Ruiz, S., Severy, G., Scherer, S., Scott, G., Shen, H., Shooshari, N., Slason, I., Sodergren, E., Sonalike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Syarik, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, C., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

TITLE

Unpublished

JOURNAL

Direct Submission

2 (bases 1 to 173224)
Submitted (08-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Genome Center

COMMENT

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project information

Center project name: GJMK

Center clone name: CH230-17N23

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 110803 bases at least Q40

Consensus quality: 120190 bases at least Q30

Consensus quality: 127171 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 77 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1	1639:	contig of 1639 bp in length
*	1739:	gap of unknown length
*	1740:	contig of 1162 bp in length
*	2902	3001: gap of unknown length
*	3002	4294: contig of 1293 bp in length
*	4295	4394: gap of unknown length
*	4395	5609: contig of 1215 bp in length
*	5610	5709: gap of unknown length
*	5710	7058: contig of 1349 bp in length
*	7059	7158: gap of unknown length
*	7159	8627: contig of 1469 bp in length
*	8628	8727: gap of unknown length
*	8728	10289: contig of 1562 bp in length
*	10290	10389: gap of unknown length
*	10390	11570: contig of 1181 bp in length
*	11571	11670: gap of unknown length
*	11671	13440: contig of 1770 bp in length
*	13441	13540: gap of unknown length
*	13541	14876: contig of 1336 bp in length
*	14877	14976: gap of unknown length
*	14977	16738: contig of 1762 bp in length
*	16739	16839: gap of unknown length
*	16839	18011: contig of 1173 bp in length
*	18012	18111: gap of unknown length
*	18112	19712: contig of 1601 bp in length
*	19713	19812: gap of unknown length
*	19813	21882: contig of 2070 bp in length
*	21883	21883: gap of unknown length
*	21883	23865: gap of unknown length
*	23865	23966: gap of unknown length
*	23967	25663: contig of 1697 bp in length
*	25664	25763: gap of unknown length
*	25764	27201: contig of 1438 bp in length
*	27201	27301: gap of unknown length
*	27302	28420: contig of 1119 bp in length
*	28421	28520: gap of unknown length
*	28521	29588: contig of 1068 bp in length
*	29589	29688: gap of unknown length
*	29689	31106: contig of 1418 bp in length
*	31107	31206: gap of unknown length
*	31207	32829: gap of 1623 bp in length
*	32830	32929: gap of unknown length
*	32930	34640: contig of 1711 bp in length
*	34641	34740: gap of unknown length
*	34741	36598: contig of 1858 bp in length
*	36599	36698: gap of unknown length
*	36699	38197: contig of 1499 bp in length
*	38198	38297: gap of unknown length
*	38298	39479: contig of 1182 bp in length
*	39480	39579: gap of unknown length
*	39580	41047: contig of 1468 bp in length
*	41048	41147: gap of unknown length
*	41148	42801: contig of 1654 bp in length
*	42802	42901: gap of unknown length
*	42902	44190: contig of 1289 bp in length
*	44191	44290: gap of unknown length
*	44291	46297: contig of 2007 bp in length
*	46298	46397: gap of unknown length
*	46398	48473: contig of 2076 bp in length
*	48474	48573: gap of unknown length
*	48574	49797: contig of 1224 bp in length
*	49798	49897: gap of unknown length
*	49898	50906: contig of 1009 bp in length
*	50907	51006: gap of unknown length
*	51007	52842: contig of 1836 bp in length

Query Match	Best Local Similarity	Score	Length	DB 2: Length
Matches 769; Conservative	57.6%; Pred. No. 1.le-18;	0; Mismatches 442; Indels 123; Gaps 21;		
QY 182	ACGAAGCAGCAGTATCACTTTTGTACGATATTGCTTGAAGTTTGGGAACCTTCG 241			
Db 122214	ATGCAAGGCGCAGTATCACTTCTGTACGATGCTTGAAGTTTCTTCAAGATCTTCG 122273			
QY 242	ACTTGGATTGAAGAAAGCTCTGTGCTCTCACTGTGAATTAACCAAGGGTTGCAC 301			
Db 122274	GCTTGAATTTAAGGAAGACTTC-CACGGCTGTCACTGTGAGACTGCCAAGCAGCTGGAGC 122332			
QY 302	CTCCTCATAAAGAACATGTTGCACATGTGCTGAAGGGCTTGGTATGATCATATCTGCT 361			
Db 122333	-----TGGGCTGTCTGGAATTG-TCTGCTGGCCGACGACATTTGCT 122371			
QY 362	TTCTTGGTTATCACTTATATTTCTTTCTTAAAGCTCCTGGAAGGCAATATCATTTGGC 421			
Db 122372	TTCTGCTTCTTCTGAATAATCTG-----AGGCGACAGACTCTGGGC 122413			
QY 422	TTGGGGTATCAGTGTATTACTATTGATCTTGTCTAGGCAATATCAAAATATCTCCACA 481			
Db 122414	TTCGAGTAAACCCCTTTCAGTTGATCTTCTG- - - - -ATATCAAAAATATCCCT-CCATG 122467			

QY	482	TTTTCCAGTGAACAGATGTTACATATAAACGATTTGACGCTTGCTGATTTGGTTGAAGGGG	541
Db	122468	CTTTCACATGAAACCTAAACCTTACGTGAACAACCTCAGCTTGCTGTTGGTTGTGGTA	122527
QY	542	TTACAGAGCCCAATPAAAGATTATAAAATATATCATTAAGATTTTATTTGAAGCGTGC	601
Db	122528	TTACAGAGTTTAATATAAAGACTT-AAATGTTCTGTTAAGATTTTATTTGAAGAGGTGC	122586
QY	602	TGAGAGAGCTGAGGATTTCCAGAGCTTTGTAAGTTCTTATTCCTGGAGAACATAAGCC	661
Db	122587	CAAAAGCCTCAGACAAATTTCCAAACCTTACAAAGC-CTGACTCTGGAGAACATAGGGGC	122645
QY	662	AATAATCATGACCTCTCCAGCATTTTTAAGACAGATGCTATATCATATGTTCTTAACTA	721
Db	122646	AATAAAC-TGCAAGGCTCACAGCTGACCCCTTGTGGGAGTTAGCAGACATTTTATCT-	122703
QY	722	GAGCGTGTACTTTTGTGGGCAATTTGAATAACCCAGATTTAAAAAGATCCAGTTAGGATG	781
Db	122704	GTGTATATACCTCCAGCTGTGTACATGACCTACTC-----AAAGGACTAATATAGATG	122755
QY	782	GACTAACTTTGGACACAATAATGGCTCTCCATTTCCATATTTGATA-----	827
Db	122759	GACTAACTTTGATATPAAATGCGGCTTTGTGCTTATACCTCCAGTCACAGTCACAGTGC	122818
QY	828	-----CTGCTGCCCTTCTACAGCTGCTAGACCAAGAC	859
Db	122819	TGGGATTTTCGAGAGAGAGAGAGAGAGAGCTGTGCTGTGAGCGGACATGAGTCTTGT	122878
QY	860	CTGTTGGTCTGGGAAGCATTTTCATGGATPAGGGAGACTCCT-----CTCGGT	906
Db	122879	CTGCTGCCCTGGGAAGCCCCCTGTGTAGGTGAATGCTTCCCTCTTACCTGACACACTGG	122933
QY	907	GNACATCCAAAACATAAATAGATGTTTATPAGAAAGCCAGAGAGACCTTTGGCAT	966
Db	122939	TGAACATCCCAAACTPAAATCCACATCCATCCCTAAAGTC-AGTGGAAATGTTGGCAG	122996
QY	967	GCGTAGTCTTCTCTATCCCACTAAC----ACTTAACATATTACTAGTCTGCTTGT	1023
Db	122997	ACCCCAACTACACTTCTTCTCTCTCCGCCCGAGTCAATGTACTCTCCACTGCTTAT	13056
QY	1024	TAAAGCAAGTATTCCTTAACCTGCTCTCTCTGCTCTTGGCCCTTAAGCTAACATPAA	1083
Db	123057	CAGAAAGCAATGATCATATACACTCTCTCTATATCTGT-AACTAGCCCTACATTTAAT	123115
QY	1084	GTTTGATATAGCATTATATATATTAATTCGTAGTCATTCATGTATCTCATGTGTGATG	1143
Db	123116	TTAATACATGATATGTGCTGTGAATTTGAGTCATCTTGTGTAACCTGT--TGTCGAAG	123172
QY	1144	TATTTTCAACTAGATCTATGATATGTTTTTTCCAGAGTTCATTAATCATTTAT	1203
Db	123174	TTTTTTCTTTTAACTAAGATCTAATAAAAAAAAGCTTTTTCCTCTCTGAGGACCA	123233
QY	1204	TGCTTACTCTTCCACCTGTGTGAACAATTTGAAGACTGAGATTGGGAACCCCAATTTG	1263
Db	123234	TTAAGCTATTTCTTCTCTATAGAGATATTTAGATGCTGGACAT- GAACCTGATTTTG	123292
QY	1264	GAAAACAGATTCATATGATGAAATGAAATCTTCATATTCGTGTTTTGAAAGATGT	1323
Db	123293	AAGTACCAATTCATATGTTGTGAATGAAATCTTCATATGCTGTGTAAAGAAA-TTCT	123355
QY	1324	GGCATTATTAACAGTATTTTATTAATAGACTTGGCTGCTGTAACATTTAATATGATATTT	1383
Db	123352	GGACATTTGTTGATTTATCATAT--CAGATTCCTGCTCAGACAAT- GTAAGTATTTCT	123408
QY	1384	TGACACAAGAGTTCTGTGAGCAAGCTATATCTATATATAGCTATAAACAATAGATATG	1443
Db	123409	TGGTTGAAGGTTTCAATGATCAGACAATCTTAATATTAAGCCACAAAGAAATGATTAAG	123468
QY	1444	AGTGTGTGACGT 1457	
Db	123469	TGTATTTGTACGT 123482	

RESULT 9
AC118715/c
LOCUS AC118715 55785 bp DNA linear HTG 20-APR-2002
DEFINITION Mus musculus clone RP24-14601, LOW-PASS SEQUENCE SAMPLING.
AC118715
ACCESSION AC118715.1 GI:20219209
VERSION
KEYWORDS HTG; HTGS; PHASE0.
SOURCE
MUS MUSCULUS.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 55785)
Birten,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP24-14601
Unpublished
2 (bases 1 to 55785)
Birten,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArrelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Glade,S., Goid,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Huime,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,U., Margulis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Melidim,J., Meneus,L.,
Milnova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Nobdu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunhkhang,P., Plerre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppach,R.,
Seaman,S., Sevelly,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wymann,D., Ye,W.J., Young,G.,
Zaloun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 125107
Center clone name: 146_O_1

* NOTE: This record contains 70 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 708: contig of 708 bp in length
* 709 808: gap of 100 bp
* 809 1517: contig of 709 bp in length
* 1518 1617: gap of 100 bp
* 1618 2331: contig of 714 bp in length
* 2332 2431: gap of 100 bp
* 2432 3131: contig of 700 bp in length
* 3132 3231: gap of 100 bp

3232 3931: contig of 700 bp in length
3932 4031: gap of 100 bp
4032 4749: contig of 718 bp in length
4750 4849: gap of 100 bp
4850 5555: contig of 706 bp in length
5556 5655: gap of 100 bp
5656 6360: contig of 705 bp in length
6361 6460: gap of 100 bp
6461 7156: contig of 696 bp in length
7157 7256: gap of 100 bp
7257 7942: contig of 686 bp in length
7943 8042: gap of 100 bp
8043 8734: contig of 692 bp in length
8735 8834: gap of 100 bp
8835 9525: contig of 691 bp in length
9526 9625: gap of 100 bp
9626 10318: contig of 693 bp in length
10319 10418: gap of 100 bp
10419 11147: contig of 729 bp in length
11148 11247: gap of 100 bp
11248 11959: contig of 712 bp in length
11960 12059: gap of 100 bp
12060 12765: contig of 706 bp in length
12766 12865: gap of 100 bp
12866 13577: contig of 712 bp in length
13578 13677: gap of 100 bp
13678 14375: contig of 698 bp in length
14376 14475: gap of 100 bp
14476 15180: contig of 705 bp in length
15181 15280: gap of 100 bp
15281 15972: contig of 692 bp in length
15973 16072: gap of 100 bp
16073 16737: contig of 665 bp in length
16738 16837: gap of 100 bp
16838 17539: contig of 702 bp in length
17540 17639: gap of 100 bp
17640 18366: contig of 727 bp in length
18367 18466: gap of 100 bp
18467 19171: contig of 705 bp in length
19172 19271: gap of 100 bp
19272 19959: contig of 688 bp in length
19960 20059: gap of 100 bp
20060 20746: contig of 687 bp in length
20747 20846: gap of 100 bp
20847 21560: contig of 714 bp in length
21561 21660: gap of 100 bp
21661 22349: contig of 689 bp in length
22350 22449: gap of 100 bp
22450 23146: contig of 697 bp in length
23147 23246: gap of 100 bp
23247 23935: contig of 689 bp in length
23936 24035: gap of 100 bp
24036 24737: contig of 702 bp in length
24738 24837: gap of 100 bp
24838 25559: contig of 722 bp in length
25560 25659: gap of 100 bp
25660 26363: contig of 710 bp in length
26370 26469: gap of 100 bp
26470 27175: contig of 706 bp in length
27176 27275: gap of 100 bp
27276 27971: contig of 696 bp in length
27972 28071: gap of 100 bp
28072 28780: contig of 709 bp in length
28781 28880: gap of 100 bp
28881 29586: contig of 706 bp in length
29587 29686: gap of 100 bp
29688 30392: contig of 706 bp in length
30393 30492: gap of 100 bp
30493 31166: contig of 674 bp in length
31167 31266: gap of 100 bp
31267 31965: contig of 697 bp in length
31964 32063: gap of 100 bp
32064 32770: contig of 707 bp in length

```

* 32771 32870: gap of 100 bp
* 32871 33593: contig of 723 bp in length
* 33594 33693: gap of 100 bp
* 33694 34377: contig of 684 bp in length
* 34378 34477: gap of 100 bp
* 34478 35182: contig of 705 bp in length
* 35183 35282: gap of 100 bp
* 35283 35968: contig of 686 bp in length
* 35969 36068: gap of 100 bp
* 36069 36763: contig of 695 bp in length
* 36764 36863: gap of 100 bp
* 36864 37548: contig of 685 bp in length
* 37549 38326: contig of 678 bp in length
* 38327 38426: gap of 100 bp
* 38427 39114: contig of 688 bp in length
* 39115 39214: gap of 100 bp
* 39215 39871: contig of 657 bp in length
* 39872 39971: gap of 100 bp
* 39972 40658: contig of 687 bp in length
* 40659 40758: gap of 100 bp
* 40759 41465: contig of 707 bp in length
* 41466 41565: gap of 100 bp
* 41566 42253: contig of 688 bp in length
* 42254 42353: gap of 100 bp
* 42354 43056: contig of 703 bp in length
* 43057 43156: gap of 100 bp
* 43157 43836: contig of 680 bp in length
* 43837 43936: gap of 100 bp
* 43937 44640: contig of 704 bp in length
* 44641 44740: gap of 100 bp
* 44741 45438: contig of 699 bp in length
* 45440 45539: gap of 100 bp
* 45540 46231: contig of 692 bp in length
* 46232 46331: gap of 100 bp
* 46332 47034: contig of 703 bp in length
* 47035 47134: gap of 100 bp
* 47135 47858: contig of 725 bp in length
* 47860 47959: gap of 100 bp
* 47960 48654: contig of 695 bp in length
* 48655 48754: gap of 100 bp
* 48755 49450: contig of 696 bp in length
* 49451 49550: gap of 100 bp
* 49551 50246: contig of 696 bp in length
* 50247 50346: gap of 100 bp
* 50347 51030: contig of 684 bp in length
* 51031 51130: gap of 100 bp
* 51131 51820: contig of 690 bp in length
* 51821 51920: gap of 100 bp
* 51921 52585: contig of 665 bp in length
* 52586 52685: gap of 100 bp
* 52686 53389: contig of 704 bp in length
* 53390 53489: gap of 100 bp
* 53490 54188: contig of 699 bp in length
* 54189 54288: gap of 100 bp
* 54289 54977: contig of 689 bp in length
* 54978 55077: gap of 100 bp

```

Query Match Best Local Similarity 5.8%; Score 106.6; DB 2; Length 55785; Pred. No. 3e-08; Matches 308; Conservative 0; Mismatches 134; Indels 41; Gaps 8;

```

QY 186 AGGAGCAGTACCTTTGTTACATATGCTTGAGGTTCTGGAAACTCTGACTT 245
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44434 AGGAGCAGTACCTTTGTTACATATGCTTGAGGTTCTGGAAACTCTGACTT 44375
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 246 TGGATTAGAAAGACTCTGTTGCTCTCACTTGAATTACCAAGTGGTTGACCTCC 305
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44374 CGATTAGAGAGACTGTG-CGCTGTGCTCCCTCGAGGTACCGACAGCTTGAGGC--- 44319
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 306 TCATTAAGAACATGTTGCTGCTGCTGAGAGGCTTTCTATGCTATACATATCTGCTTCT 365
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44318 -----TGAGCGCTGCTGAAGG-TCGCGGCGCGTGCAGCTGCTGCTTCT 44276
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

QY 366 TGGTTATGATTTATTTCTTTCTAAAGCTCCCTGAAGGCAATATCATTTGGCTGG 425
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44275 GATTTT-----TCTCTGAAAGTCCCTGAGGTAAGCACTACTGCGACAG 44230
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 426 GGTGATCAGTGTCTTACTTATTTGATCTTGTGCTAGGCAATATCAATCACTCCACATTTT 485
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44229 AGTGAACGTGTTCCAC---TTGATCTTTCTGACAAAGAGCAAAATACCC-CCATGCTT 44174
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 486 CCAGTGAAACAGATGTTACATTAAGCAATGACGTGCTGCTATTTGGTTAAGGATTAAC 545
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44173 CTACGAAACGGAAGTGCATGAACAACCTCCGCTGCTGTGTGTGTGTATAC 44114
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 546 AGAGCCATTAAGATTTAAATATATCATTAAGATTTATTTGGAAGGTGGCTGGA 605
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44113 AGAGCTTAATTAAGACTT-AGATGTGTTGATTAAGATTTATTTGGAAGGTGGTAAA 44055
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 606 GAGAGCTGAGATTTCCAGACACTTGTGAATTTCTATTTGGAAGCAATAGCCATA 665
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 44054 AGCCTGACAAATTTCTTAACACTTCAAG-CCCTACTGTGGGGAACATGGGGCAGTA 43996
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 666 ATC 668
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 43995 AAC 43993
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 10
AC073269 137889 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-436F9 from 7, complete sequence.
DEFINITION AC073269
AC073269.7 GI:14269824
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 137889)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 137889)
AUTHORS VanBrunt, A., Abbott, A., and McElliott, M.
TITLE The sequence of Homo sapiens BAC clone RP11-436F9
JOURNAL Unpublished (2001)
3 (bases 1 to 137889)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 137889)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 137889)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 137889)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 137889)
AUTHORS Waterston, R.
TITLE Direct Submission

```

JOURNAL
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 31, 2001 this sequence version replaced g1:13794256.

COMMENT

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0436P09

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
<http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send
<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The Rpt-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is Rp11-29C24, 2000 bp overlap; the
clone sequenced to the right is CTD-2016H9, 200 bp overlap. Actual
start of this clone is at base position 1 of Rp11-436F9; actual end
is at base position 92636 of CTD-2016H9.

Polymorphisms have been identified between AC069313, AC021448 and
AC073269.

FEATURES

SOURCE

Location/Qualifiers
1. 137889
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="Rp11-436F9"
/clone_lib="Rpt-11"
repeat_region
453..502
/rpt_family="AT-rich"
774..1051
/rpt_family="Alu"
repeat_region
1063..1316
/rpt_family="L1"
1351..1469
/rpt_family="L1"
repeat_region
2098..2256
/rpt_family="MIR"
repeat_region
2505..2753
/rpt_family="L1"

repeat_region 2971..3034
/rpt_family="L1"
repeat_region 3558..3742
/rpt_family="L2"
repeat_region 3905..4033
/rpt_family="L2"
repeat_region 4113..5347
/rpt_family="L1"
repeat_region 4854..4888
/rpt_family="CT-rich"
5341..5560
/rpt_family="Alu"
repeat_region 6804..6903
/rpt_family="L2"
repeat_region 6920..7957
/rpt_family="MER1-type"
7958..8260
/rpt_family="Alu"
repeat_region 8261..8370
/rpt_family="MER1-type"
8826..8873
/rpt_family="ERV1"
8833..8877
/rpt_family="(TG)n"
8836..9211
/rpt_family="L2"
repeat_region 9345..9387
/rpt_family="L2"
repeat_region 10487..10723
/rpt_family="MER2-type"
13743..14036
/rpt_family="Alu"
14027..14069
/rpt_family="(TAAA)n"
14540..14937
/note="match to EST AV729823 (NID:g10839244)"
14800..14872
/note="match to EST AV709400 (NID:g10726999)"
14800..14872
/note="match to EST BF966467 (NID:g12333682)"
14824..14872
/note="match to EST A1138808 (NID:g3644780) qd98g02.x1"
14918..15201
/rpt_family="ERV1"
15247..15529
/rpt_family="(TCTA)n"
15308..15546
/rpt_family="ERV1"
15566..15748
/rpt_family="(TCTA)n"
15626..15894
/rpt_family="ERV1"
15764..15802
/rpt_family="(TCTA)n"
15814..15980
/rpt_family="(TCTA)n"
15990..16035
/rpt_family="ERV1"
16102..16395
/rpt_family="MALR"
16405..17052
/note="match to EST BF966467 (NID:g12333682)"
16872..17339
/note="match to EST AV709400 (NID:g10726999)"
16872..17204
/note="match to EST A1138808 (NID:g3644780) qd98g02.x1"
17326..17347
/rpt_family="AT-rich"
17664..17970
/rpt_family="Alu"
17946..17977
/rpt_family="A-rich"
18654..18690

repeat_region /rpt_family="MIR"
18709..18849
repeat_region /rpt_family="L2"
19230..19400
repeat_region /rpt_family="TATATC)n"
19476..19668
repeat_region /rpt_family="Alu"
19850..20013
repeat_region /rpt_family="L2"

Query Match 5.3%; Score 97.6; DB 9; Length 137889;
Best Local Similarity 47.4%; Pred. No. 6.4e-07;
Matches 361; Conservative 0; Mismatches 394; Indels 7; Gaps 2;

1073 TAACTAATAAGTTGATATAGGCAATTAATTAATTCGAGTCATCATGATCTCT 1132
1133 CATGTTGATGATGTTTTCACAACTAGATGATGATGATGATGATGATGATGAT 1192
63447 ATTTTAT 63506
1193 AATCATTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1252
63507 AAT 63566
1253 ACCCAATTTGGAAACCAAGTTCATGATGATGATGATGATGATGATGATGAT 1312
63567 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 63626
1313 TGAAGAATGAGGCAATTTATATATATATATATATATATATATATATATAT 1372
63627 TTAACAT 63682
1373 TAGGATATTTTGGACAGAGGTTCTGTCACACATTTACCTAATTTATACCT 1432
63683 TATTTTATTTTATATATATATATATATATATATATATATATATATATAT 63742
1433 CAATATGATGAGGCTTGTGATGATGATGATGATGATGATGATGATGATGAT 1492
63743 TAT 63802
1493 AGAATAATGTTAATATATATATATATATATATATATATATATATATATAT 1552
63803 TTTTAAATGTTCAATATATATATATATATATATATATATATATATATAT 63862
1553 TGCGTGAATGAT 1612
63863 AAT 63922
1613 TATATGCAAAAT 1669
63923 TAT 63982
1670 CATATGAAAT 1729
63983 TTAAT 64042
1730 CTAT 1789
64043 AT 64102
1790 CAGAGAAATCTGATGATATATATATATATATATATATATATATATATAT 1831
64103 CAT 64144

RESULT 11
AL603966/c 103737 bp DNA 1linear PRI 30-JAN-2002
LOCUS Human DNA sequence from clone RP11-164N7 on chromosome 10, complete
DEFINITION sequence.
ACCESSION AL603966

VERSION AL603966.9 GI:18476754
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 103737)
AUTHORS Bird,C.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humuery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced gi:16605768.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em.; EMBL; Sw.;
SWISSPROT; Tr.; TrEMBL; Mp.; WORMREP; Information on the WORMREP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormrep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/MGP/chr10
RP11-164N7 is from the library RPCT-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone
RP11-164N7. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP11-164N7 is at 1 in this sequence. The
true left end of clone RP11-13E1 is at 101738 in this sequence.
FEATURES
source
1..103737
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-164N7"
/clone_11d="RPCT-11.1"
/clone_11b="RPCT-11.1"

BASE COUNT 31673 a 22545 c 21304 g 28215 t
ORIGIN

Query Match 5.2%; Score 95; DB 9; Length 103737;
Best Local Similarity 95.1%; Pred. No. 1.8e-06;
Matches 98; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

28 TTTATGCTTTAGAACTTGAATCTGCTGCTGATGATGATGATGATGATGATGATG 87
96407 TTAAGGCTTTAGAACTTGAATCTGCTGCTGATGATGATGATGATGATGAT 96348
88 AACAGAGAAAAAACAGATACCGAGATTTCTTCATTTCAAC 130
96347 AACAGAGAAAAAACAGATACCGAGATTTCTTCATTTCAAC 96305

RESULT 12
AL450334/c 171930 bp DNA 1linear PRI 21-JUN-2002
LOCUS Human DNA sequence from clone RP11-556L1 on chromosome 10, complete
DEFINITION sequence.
ACCESSION AL450334

3	AC005504	104992 bp	DNA	linear	HMG 01-APR-1999
2	AC005504/c				
1	AC005504				
	AC005504.3				
	GR:4558584				
	HMG: HNGS_PHASE1.				
	Plasmodium falci-				
	parum.				
	Plasmodium falci-				
	parum				
	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
	1 (bases 1 to 104992)				
	Hyman, R.W., Fung, E.L., Qin, F., Tanaki, T., Kurd, O.B., Conway, A.B.				
	and Davis, R.W.				
	Plasmodium falci-parum 3D7 chromosome 12				
	Unpublished				
	2 (bases 1 to 104992)				
	Hyman, R.W., Qin, F., Fung, E.L., Conway, A.B. and Davis, R.W.				
	Direct Submission				
	Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology				
	Center, Stanford University, 855 California Avenue, Palo Alto, CA				
	9404, USA				
	On Apr 2, 1999 this sequence version replaced gi:4337172.				
	* NOTE: This is a 'working draft' sequence. It currently				
	* consists of 3 contigs. The true order of the pieces				
	* is not known and their order in this sequence record is				
	* arbitrary. Gaps between the contigs are represented as				
	* runs of N, but the exact sizes of the gaps are unknown.				
	* This record will be updated with the finished sequence				
	* as soon as it is available and the accession number will				
	* be preserved.				
	1 58642: contig of 58642 bp in length				
	58643 58842: gap of unknown length				
	58843 91011: contig of 32169 bp in length				
	91012 91211: gap of unknown length				
	91212 104992: contig of 13781 bp in length.				
	Location/Qualifiers				
	SATRES				

	source	1. .104992 /organism="Plasmodium falciparum" /db_xref="taxon:5833" /chromosome="12"	
BASE COUNT	44286 a 9326 c 9564 g 4141 t	405 others	
ORIGIN			
Query Match	4.7% ; Score 85.2; DB 2;	Length 104992;	
Best Local Similarity	47.3%; Pred. No. 6,7e-05;		
Matches 366; Conservative 0; Mismatches 393; Indels 15; Gaps 3;			
OY	1067 TTTAGCTAACTAATTAAGTGTGATATAGCGCATATTATATATATTCGAGTCATTCATGCT	1126	
Db	73856 TTTCACATTTTAAATAAATAAATAAACCCCTTTTATTAATACCTATTAAATTAATTAATA	7379	
OY	1127 ATCTCCTATGGTGGATGATTTTTCCAACATGAAGTCATGATGATCTTTTTTCCAGAGT	1166	
Db	73796 ATAATAATTAATTAATGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	7373	
OY	1187 CCATTAAATCAATTTATTTCCCTTTACTTTCCTCCACCCTCGTGAAACATT--TAGAACTG	1243	
Db	73736 AATATAATATATATATATATATCTTAATTAATTAATTAATTTTAAATTTATTAATTAAT	7367	
OY	1244 GATTGGGAACCCCAATTTTGGAAAACAGATTCATAGCATGAAAAAGAAACTGCCATA	1303	
Db	73676 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	7361	
OY	1304 TTCCTGTTTGGAAAAGAAGTGGCCCATTAATTAACAGTAATTTTATTAATGAAGCTTGGCCCTG	1363	
Db	73616 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	7355	
OY	1364 TACAATTAAATAGTATATTTTGGACAAGAGCTCTGGACAGCTATCTAATTAATA	1423	
Db	73556 AATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	7349	
OY	1424 GCATTAATAACATATGATATGAGTCTTTGACAGTTTAACATCAATGAGATCAGA-AATAT	1482	
Db	73496 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT	7343	
OY	1483 CTATGATATGAGAAATGTTTAATATCACTATTAATCTTGAAATTTCTGAAGGCTTAT	1542	
Db	73436 TAATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	7337	
OY	1543 TTTCCTCTTTGGCGTGAATGATATTTGAAATGGTTGAATTAATTAATTCATG	1602	
Db	73376 ATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTA	7331	
OY	1603 TAAATAATTAATTAATGCCTAAATAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAAT	1654	
Db	73316 TAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	7325	
OY	1655 ---TCGTTCATGTTGTCATATGCAATATATATCTGCTATTAATTCATTTGATTAATGG	1711	
Db	73356 AATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	7319	
OY	1712 CCTTAATTAATTTGATGCTCAATTTATTTCTGCTGCTGTAATAATATATAGCTATA	1771	
Db	73196 AAATTAATTAATTAATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTA	7313	
OY	1772 ATGATTAACAATTTTCCTCAGAGAAATTCATGCTATTAATTAATAAATAATTT	1825	
Db	73136 TAATTAATGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	73083	
RESULT 15			
AC004157/c	AC004157	169546 bp	DNA linear HTG 12-AUG-2000
LOCUS	Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN		
DEFINITION	PROGRESS ***, 2 unordered pieces.		
ACCESSION	AC004157		
VERSION	AC004157.8 GI:9797712		
KEYWORDS	HTG; HTGS_PHASE1.		
SOURCE	Plasmodium falciparum.		

